DOCUMENT RESUME

ED 452 225 TM 032 510

AUTHOR	Glass, Gene V., Ed.
TITLE	Education Policy Analysis Archives, 2001.
ISSN	ISSN-1068-2341
PUB DATE	2001-00-00
NOTE	322p.; "Education Policy Analysis Archives" is an electronic-only journal (see http://olam.ed.asu.edu/epaa) covered on an article-by-article basis in "Current Index to Journals in Education" (CIJE). For the 11 articles in this part of Volume 9, see TM 523 384-394 (in CIJE).
PUB TYPE	Collected Works - Serials (022)
JOURNAL CIT	Education Policy Analysis Archives; v9 n1-11 2001
EDRS PRICE	MF01/PC13 Plus Postage.
DESCRIPTORS	*Dropouts; *Educational Research; Educational Technology; Educational Vouchers; Elementary Secondary Education; English (Second Language); Higher Education; Internet; Puerto Ricans; *School Choice; School Segregation; *Student Attitudes; *Teacher Education; *Test Use; Values

ABSTRACT

This document consists of downloaded copies of 11 papers published in the electronic journal "Education Policy Analysis Archives" during 2001: (1) "School Segregation of Children Who Migrate to the United States from Puerto Rico" (Luis M. Laosa); (2) "Testing Times: A School Case Study" (Ivor Goodson and Martha Foote); (3) "Impact of U.S. Overseas Schools in Latin America on Political and Civic Values Formation" (John J. Ketterer and George E. Marsh, II); (4) "Factors Influencing GED and Diploma Attainment of High School Dropouts" (Jeffrey C. Wayman); (5) "How the Internet Will Help Large-Scale Assessment Reinvent Itself" (Randy Elliot Bennett); (6) "Teacher Test Accountability: From Alabama to Massachusetts" (Larry H. Ludlow); (7) "Critique of 'An Evaluation of the Florida A-Plus Accountability and School Choice Program" (Gregory Camilli and Katrina Buckley); (8) "The Effects of Vouchers on School Improvement: Another Look at the Florida Data" (Haggai Kupermintz); (9) "The Academic Journal: Has It a Future?" (Gaby Weiner); (10) "Japanese EFL Teachers' Perceptions of Communicative, Audiolingual and Yakudoku Activities: The Plan versus the Reality" (Greta Gorsuch); and (11) "Constructing Outcomes in Teacher Education: Policy, Practice and Pitfalls" (Marilyn Cochran-Smith). (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

VOLUME 9 - 2001

ISSN 1068-2341

ED 452 225

Education Policy Analysis Archives

(Articles 1 thru 11)

TM032510

editor:
Gene V. Glass
Arizona State University

BEST COPY AVAILABLE

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL HAS
BEEN GRANTED BY

Gene Glass

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

● Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

EDUCATION POLICY ANALYSIS ARCHIVES

Volume 9, 2001 **Articles 1-10**

Article		Pages
1	Luis M. Laosa: Segregation of Children Who Migrate to the U.S. From Puerto Rico	57
2	Ivor Goodson & Martha Foote: Testing Times: A School Case Study	12
3	John J. Ketterer & George E. Marsh: Impact of U.S. Overseas Schools in Latin America on Values	11
4	Jeffrey C. Wayman: Factors Influencing Dropouts' GED & Diploma Attainment	23
5	Randy E. Bennett: How the Internet Will Help Large-Scale Assessment Reinvent Itself	30
6	Larry H. Ludlow: Teacher Test Accountability: From Alabama to Massachusetts	27
7	Gregory Camilli & Katrina Bulkley: Critique of "An Evaluation of the Florida A-Plus Accountability Program"	20
8	Haggai Kupermintz: Effects of Vouchers: Another Look at the Florida Data	15
9	Gaby Weiner: The Academic Journal: Has it a Future?	23
10	Greta Gorsuch: Japanese EFL Teachers' Perceptions of Communicative, Audiolingual and Yakudoku Activities The Plan Versus the Reality	33
11	Marilyn Cochran-Smith: Constructing Outcomes in Teacher Education: Policy, Practice and Pitfalls	68

This article has been retrieved 1106 times since January 1, 2001

other vols. | abstracts | editors | board | submit | comment | subscribe | search

Education Policy Analysis Archives

Volume 9 Number 1 **January 1, 2001** **ISSN 1068-2341**

A peer-reviewed scholarly journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2001, the **EDUCATION POLICY ANALYSIS ARCHIVES**.
Permission is hereby granted to copy any article
if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education.*

School Segregation of Children Who Migrate to the United States From Puerto Rico

Luis M. Laosa
Educational Testing Service
Princeton, New Jersey

Abstract

This study examined patterns of school segregation (ethnic/racial, linguistic, and socioeconomic) and other ecological characteristics of the schools that preadolescent children who migrate from Puerto Rico to the United States (New Jersey) attend in this country during the first two years following their arrival ($N = 89$ schools). The data show that Hispanics/Latinos are the majority of the student body in 43% of the schools; African Americans, in 30% of the schools; and European Americans, in 12% of the schools. Native speakers of Spanish are the majority of the student body in 29% of the schools. Approximately one half of the schools are in economically depressed, highly urbanized areas. Although the schools are on average large, 44% of them enroll above capacity. In most schools the majority of the student body is from economically impoverished families with low levels of parental education. There are, however, wide differences among the schools on each of these variables. Correlations show that the higher a student body's proportion of Hispanics/Latinos or native speakers of Spanish, the higher is the student body's proportion of pupils from economically impoverished households with low levels of parental education, and the higher the school's likelihood of being crowded and of being located in a poor inner-city area. Similarly, the higher a student body's proportion of African Americans, the higher is the student body's proportion of pupils from low-income families, and the higher the school's likelihood of being in a poor inner-city area. The findings are discussed with regard to implications for policy and hypotheses in need of research concerning possible consequences of school segregation for students' academic, linguistic, social, and emotional development. Also presented is a historical overview, to the present, and discussion of U.S. policies and judicial decisions concerning school segregation, with particular reference to segregation of Hispanics/Latinos.

Introduction

Schools are social institutions ecologically niched in individual communities that are in turn embedded in larger, layered systems. Thus, each school functions as part of a social, cultural, political, and economic environment. What each school is like will be determined in part by this ecology. In the United States, vast ecological differences exist among schools. This subject raises a broad range of issues, including questions about resource allocation, the distribution of power in society, and educational ideologies (see, e.g., Barton, Coley, & Goertz, 1991; Cobb & Glass, 1999; Kennedy, Jung, & Orland, 1986; Laosa, 1984; Minuchin & Shapiro, 1983; Orland, 1994; Puma, Jones, Rock, & Fernandez, 1993; Rutter, Maughan, Mortimore, & Ouston, 1979; Southern Education Foundation, 1995; U.S. Department of Education, 1993b, 1996, 1997). The subject also raises serious questions about the role of schools in creating or maintaining socioeconomic stratification and ethnolinguistic isolation. These considerations bear especially on children from immigrant and other ethnocultural and linguistic minority groups. For many of these children, the school is the first—and perhaps the only—influential point of direct experience with a "mainstream" socializing institution.

In recent years, many reformers and critics of the U.S. system of education have stressed the importance of academic standards, accountability, and student assessment, whereas less attention has been given to other critical dimensions of the ecology of schools. In contrast, ecological approaches stress the context of events and encourage the search for recurrent patterns that describe the characteristics of a system. From this perspective, no unit is considered separable from the system as a whole (see, e.g., Bronfenbrenner, 1979, 1995; Laosa, 1999; Laosa & Henderson, 1991; Minuchin & Shapiro, 1983).

The study reported here examines specific dimensions of the ecology of schools, focusing particularly on the schools attended by children who migrate to the United States from Puerto Rico. Puerto Ricans are the largest Hispanic/Latino population in the Northeast of the United States (Pérez & Martínez, 1993; U.S. Bureau of the Census, 1992, 1996). Because of the special sociopolitical relationship between the two countries, (Note 1) making Puerto Ricans U.S. citizens by birth, Puerto Ricans are not, technically speaking, "immigrants" in the same sense as are entrants from nations under the jurisdiction of U.S. immigration laws. Yet, Puerto Ricans who migrate to the United States possess all the characteristics of an immigrant group, including a distinct culture and a different language—Spanish. Puerto Ricans in this country, as a group, fare worse than does the U.S. Hispanic/Latino population as a whole—and far less well than the U.S. non-Hispanic/non-Latino White population—on many socioeconomic characteristics, including varied measures of employment, income, and academic achievement (Pérez & Martínez, 1993; U.S. Bureau of the Census, 1994a, b, 1996). The study reported here is guided by the view that in order to gain a better understanding of children's development and adaptation, one must first describe the attributes of the human environments they face.

Particularl y in the United States, critical ecological attributes of schools include the student body's ethnic/racial, linguistic, and socioeconomic composition. National trends show that school segregation of African American children declined dramatically from the mid-1960s through the early 1970s; it then remained to a large extent stable until the late 1980s when, in a reversal of this trend, it began to rise. In sharp contrast, school segregation of Hispanic/Latino children has continued to increase steadily since at least the mid-1960s, when national data on the subject were first collected (Orfield, 1993; Orfield, Bachmeier, James, & Eitle, 1997; Orfield & Yun, 1999; U.S. Department of Education, 1995).

The level of school segregation for Hispanic/Latino children is high across the country; it is highest for the substantially Puerto Rican population of the Northeast, although it is rapidly rising in other regions with significant concentrations of Hispanics/Latinos. African Americans, too, face the highest segregation levels in the Northeast, although they encounter rising levels in other regions because of resegregation trends (Orfield, 1993; Orfield et al., 1997; Orfield & Yun, 1999). The highest levels of school segregation occur in urban areas, particularly in the inner core of cities.

Of greatest concern, national data further show a relationship of ethnic/racial segregation to poverty: Both Hispanic/Latino and African American children are much more likely than European American children to find themselves in schools of concentrated poverty (Orfield, 1993; Orfield, Eaton, & the Harvard Project on School Desegregation, 1996; Orfield et al., 1997; Orfield & Yun, 1999; Orland, 1994; Puma et al., 1993; U.S. Department of Education, 1993b, 1996, 1997). Although *socioeconomic status* (SES) typically refers to the background of individuals, a growing body of research suggests that the SES of a child's school may be as critical an influence on the child's academic achievement as is the SES of the child. Individual differences in children's academic performance have been shown to correlate not only with the children's household SES but also with the SES of their schools' student bodies (Kennedy et al., 1986; Orland, 1994; Puma et al., 1993; U.S. Department of Education, 1993b, 1996, 1997; U.S. General Accounting Office, 1992). For example, on the basis of a nationally representative sample of U.S. elementary students, Kennedy et al. (1986) and Orland (1994) concluded that the higher a school's concentration of economically impoverished students, the higher tends to be the incidence of low academic achievers. This relationship held even after statistically controlling for demographic characteristics of the individual students and of their families (Kennedy et al., 1986, chap. 2; Myers, 1985; Orland, 1994). Other studies lead to similar conclusions (e.g., Puma et al., 1993; U.S. Department of Education, 1993b, 1996, 1997; U.S. General Accounting Office, 1992).

Unlike previous research, the present study focuses on a specific Hispanic/Latino population and follows it longitudinally, centering on a specific chronological age period and a specific stage in the migration process. The target age is preadolescence, an age when children typically position themselves for the marked physiological and psychological changes of adolescence. Informal observations suggest that academic and psychosocial problems experienced by many Hispanic/Latino and other ethnic/racial minority students emerge during this developmental stage. The target phase of the process of migration and settlement is the first two-year span immediately following arrival in the United States, a phase when stressful demands are often placed on the individual for personal change and adaptation (Laosa, 1990, 1997, 1999).

Specifically, this study examines the following ecological attributes of the schools that preadolescents who migrate from Puerto Rico to the United States (New Jersey) attend in this country during the first two years following their arrival: the ethnic/racial, linguistic, and socioeconomic mix of the schools' student bodies; the degree of urbanness and the economic status of the neighborhoods in which the schools are located; and the schools' size and density-overcrowdedness. Also examined are the associations among these attributes. The data and analyses sought answers to the following questions concerning these schools:

- What is the ethnic/racial composition of the schools' student bodies?
- What is the linguistic composition of the schools' student bodies?
- What are the socioeconomic characteristics of the schools' student bodies?
- In what types of neighborhoods are the schools located?
- Are the schools overcrowded? What is the size of the schools?
- What, if any, are the relationships of the student body's (a) ethnic/racial composition and (b) linguistic composition to the student body's family socioeconomic characteristics? to characteristics of the school's neighborhood? to school crowdedness and school size?

Here I examine several issues pertaining to these questions; it is organized as follows: After a section that briefly notes certain sociohistorical circumstances bearing on the present relationship between the United States and Puerto Rico and on contemporary characteristics of the Puerto Rican population, the next section describes the study's research method and procedures. Next is the presentation of the data analysis results, answering each research question. An extended Discussion section summarizes conclusions from the answers to these questions and considers implications for policy and for students' academic, linguistic, social, and emotional development, identifying hypotheses in need of research; that section also includes a historical overview, to the present, and discussion of U.S. policies and judicial decisions concerning school segregation, with particular reference to segregation of Hispanics/Latinos.

Sociohistorical Context

Puerto Rico was under the colonial rule of Spain for four centuries. Spanish is the language generally spoken in Puerto Rico; it is also the language used as the medium of instruction in Puerto Rico's public schools.

The population of Puerto Rico is composed largely of the descendants of three groups: the Spanish colonizers, the original Amerindian inhabitants—the Arawak people who developed the Taíno culture—and African slaves imported by the colonizers (Mathews & Tata, 1992; Wagenheim, 1970). Sizeable minorities of the three races constitute the extremes of the skin-color spectrum, which blend in the predominant middle. Most Puerto Ricans, therefore, are generally considered "colored" by European Americans. In Puerto Rico, fuzzy lines between racial groups discourage color discrimination, although the U.S. presence and certain attitudes and practices it has brought to the island appear to have heightened the awareness of racial differences among Puerto Ricans (Rodríguez, 1991; Wagenheim, 1970). Once slavery was abolished in 1873, the law in Puerto Rico opened public places to all (Wagenheim, 1970). Thus, unlike the U.S. mainland with its *de jure* segregation, Puerto Rico did not have racially separate public facilities such as rest rooms, water fountains, or rear sections of public vehicles.

In the second half of the nineteenth century, the United States plunged into international politics and took the road to imperialism—a foreign-policy direction with far-reaching and lasting consequences. These overseas incursions brought under the nation's jurisdiction some eight million people of color in the Caribbean basin, other parts of Latin America, and the Pacific region (Lewis, 1963; Link, 1992; Morison, 1972; Woodward, 1966). (Note 2)

U.S. involvement in Puerto Rico began with the Spanish-American War, a short and relatively bloodless war that ended with the Treaty of Paris in 1898, by which Spain ceded Puerto Rico to the United States. U.S. involvement in the Caribbean region grew in the early part of the twentieth century. U.S. military bases in that area have served to protect U.S. and European interests (e.g., during World War Two) but also provide investment opportunities, often leading to the exploitation of the peoples of the Caribbean and of other parts of Latin America and hence to dependency and resentment (Carr, 1984; Lewis, 1963; Mathews & Tata, 1992; Morison, 1972).

In 1917 the U.S. Congress passed the Jones Act, which gave limited self-government to Puerto Rico and conferred U.S. citizenship collectively on its inhabitants (Carr, 1984; Wagenheim, 1970). U.S. citizens of Puerto Rico elect a representative (i.e., a "resident commissioner") to the U.S. House of Representatives, who may speak but cannot vote except in committees. These citizens are automatically involved in wars declared by the U.S. Congress and led by the U.S. President, in whose elections they cannot participate.

Although Puerto Ricans had migrated to the continental United States before the nineteenth century, only after 1900 did they begin doing so in significant numbers. Annual inflows reached their peaks during the two decades following the end of World War Two, a period when Puerto Rico's agricultural economy was radically transformed into one based on industrial production, as U.S. tax laws encouraged the establishment of new industries (Rodríguez, 1991; U.S. Commission on Civil Rights, 1976; Wagenheim, 1970). Because the number of small farms had been sharply reduced by the introduction of large-scale, single-crop corporate agribusiness, the island had virtually lost its subsistence farming system that could have enabled many families to return to individually self-supporting farming (Moore & Pachon, 1985). Numerous workers left the agricultural sector and moved into cities along the island's coast in search for jobs. Many also migrated to large metropolitan centers in the northeastern United States, responding to those areas' expanding economies and consequent demand for low-skill work, and taking advantage of the low-cost island-to-mainland passenger flights that commercial airlines then began offering (Mathews & Tata, 1992; Wagenheim, 1970). Although annual inflows are currently below the levels reached in the 1950s and 1960s, migration from Puerto Rico to the continental United

States inevitably continues, and by all indications will continue into the foreseeable future.

Method

Preparatory Demographic Studies

To inform the development of the sampling plan, a series of empirical demographic studies (e.g., Laosa, 1998) had been conducted regarding children's migratory movements between Puerto Rico and New Jersey. Those studies were necessary because the needed demographic information was not available from centralized sources. The U.S. Immigration and Naturalization Service, a source of statistics on immigration, does not monitor Puerto Rican migration because of the special U.S.-P.R. relationship. The U.S. Bureau of the Census routinely provides demographic information on the Puerto Rican stateside population and on the population of Puerto Rico but no information bearing specifically on the present investigation's more detailed focus. Similar difficulties arose with data from other agencies and organizations that provide national and state statistics.

Sample Selection

Based on those demographic studies, a sample of 241 public elementary (Note 3) schools (27 school districts) was drawn to yield a sample as representative as possible of children migrating from Puerto Rico to urban and suburban areas and small towns in the state of New Jersey. The enrollment records of each of these schools were then continually monitored during two full, consecutive academic years (i.e., two annual migration waves). All the children who transferred in from Puerto Rico (regardless of prior migration history) to the third and fourth grades (or the equivalent for ungraded programs) in these schools at any time during those two years were identified within approximately two months of their arrival. Those who met these sample-eligibility criteria and gave informed consents (self and parental) became research participants (i.e., *focal* children). Each focal child was then followed longitudinally (from the date of his or her transfer-in from Puerto Rico), regardless of destination, for two consecutive academic years. Considerable care, time, and effort were devoted to sample identification, recruitment, and longitudinal follow-up. Consequently, as reported elsewhere (Laosa, n.d.), both the participant consent rate and the sample retention rate were quite adequate with respect to scientific sampling standards; there is no reason to suspect significant sample bias.

The children who met the sample-eligibility criteria were found widely and thinly scattered across the sample schools; many of the schools received no children who met these criteria. (Note 4) The analyses reported here are based on the schools that received the focal children directly from Puerto Rico plus the schools that these children subsequently attended stateside during their respective two-year longitudinal spans (*N* = 89 schools). Almost all are New Jersey public schools because the vast majority of the focal children who transfered out of their initial receiving schools did so either to other New Jersey public schools or back to Puerto Rico.

Variables and Measures

Measurements were taken on each school that focal children attended, as described below. (Note 5)

- **Student body's ethnic/racial composition.** A student body's ethnic/racial composition is indexed by the following seven variables (a school's measurement on a variable is the percentage (Note 6) of the school's student body belonging to the corresponding ethnic/racial category): *African American* (i.e., Black), *Asian/Pacific Islander American*, *European American* (i.e., White/Caucasian), *Hispanic/Latino*, and *other ethnic/racial groups*. *Puerto Rican* and *other Hispanic/Latino* disaggregate the Hispanic/Latino category. The first, second, third, and fifth ethnic/racial categories include only non-Hispanics/non-Latinos.
- **Student body's linguistic composition.** A student body's linguistic composition is indexed by four variables (a school's measurement on a variable is the percentage (Note 7) of the school's student body belonging to the corresponding linguistic category). Three of them divide the student body by native language: *monolingual native speakers of English*, *native speakers of Spanish*, and *native speakers of other languages*. The fourth linguistic category is *limited-English-proficient/English-language learners (LEP/ELL)*; it identifies the pupils whom the school's officials formally classified as "limited-English-proficient (LEP);" also called "English-language learners (ELL)," this classification can be applied only to pupils who are not native speakers of English.
- **Student body's family socioeconomic characteristics.** To gain a deeper understanding of the construct *socioeconomic status* as it applies to the focal issues—and thus add to its relevance for policy, practice, and theory—the present study examines seven variables that respectively measure particular social, economic, and educational characteristics of the student bodies' families. Previous studies have typically included only one of these variables as a proxy index or else have combined them into a single measure of socioeconomic status or social class. Although these variables are expected to be intercorrelated, it was deemed important for the purposes of the present study to measure and analyze them individually:
    - *Unemployment level* is the percentage (Note 8) of the school's student body living in households in which the householder (Note 9) is unemployed.
    - *Public assistance dependence level* is the percentage (Note 10) of the school's student body living in households receiving public assistance (i.e., welfare).
    - A student body's average family *economic status* is measured on a 5-point scale (1 = low income;

5 = affluent).
  - A school's *fully subsidized lunch eligibility level* is the percentage (Note 11) of the student body eligible for free lunches.
  - *Partly subsidized lunch eligibility level* is the percentage (Note 12) of the student body eligible for reduced-price lunches.
  - *Subsidized lunch eligibility level (fully + partly)* is the aggregate of the last two variables (i.e., the percentage of the student body eligible for fully subsidized lunch plus the percentage eligible for partly subsidized lunch). (Note 13)
  - Finally, *maternal schooling level* is the average level of formal education attained by the student body's mothers or female guardians, measured on a 9-point scale (1 = six years of schooling or less; 9 = doctor's degree).
- **School neighborhood's urbanness and economic status.** Two variables describe the area, or neighborhood, in which the school is located: *urbanness*, a 5-point scale (1 = rural; 5 = inner core of a city), and *economic status*, also a 5-point scale (1 = low-income area; 5 = affluent area).
- **School size and crowdedness.** Four variables pertain to school size and crowdedness: A school's *enrollment size* is the total number of students enrolled in the school in late spring. *Enrollment capacity* is the number of students for which the school was built. A school's *density-overcrowdedness level* is indexed by subtracting the school's enrollment capacity from its enrollment size (thus, a higher positive value signifies denser crowdedness than does a lower positive value). The *crowdedness dichotomy* is a dichotomous variable: 1 = the school is *not* crowded (i.e., density- overcrowdedness level is zero or negative); 2 = the school is crowded (i.e., density-overcrowdedness level is greater than zero).

Data Sources

The data, including the scale ratings, were obtained directly from the schools' principals, primarily through structured questionnaires; however, when necessary the questionnaire approach was supplemented or replaced by telephone calls and by site visits in order to examine school records and to interview principals and other school staff.

Statistical Analyses

The unit of analysis is the (unweighted) individual school. The school is not weighted (i.e., by the number of focal children attending it) in the analyses, since the present focus is on the schools that focal children attend rather than on the focal children per se. (Footnote 4 shows the frequency distribution of focal children on the schools.) The analyses examine individual differences that occur among the schools on the variables. To this end, computed were the frequency distribution of the schools on each variable, its mean, standard deviation, standard error of the mean, and skewness value. Also computed were matrices of correlation coefficients. (Notes 14 & 15) For the purposes of exposition only, the frequency distribution on any variable with a very wide range is summarized in the tables or text below by collapsing the range into a suitable number of grouping intervals; however, for the purposes of computing the statistics and performing the statistical analyses, all the variables are based on the actual detailed data.

Results

The presentation of the analysis results is organized by the research questions.

1. What is the ethnic/racial composition of the schools' student bodies?

The schools attended by the focal children have, on average, a student body that is nearly one-half Hispanic/Latino, one-third African American, 17% European American, 2% Asian/Pacific Islander American, and 2% "other." Specifically, Table 1 shows that of the five broad ethnic/racial composition variables, Hispanic/Latino has the highest mean percentage (i.e., 46.5), signifying that the schools have, on average, a student body that is 46.5% Hispanic/Latino. In finer detail, this table shows that the vast majority of the Hispanic/Latino students in these schools are Puerto Rican. Indeed, the schools have, on average, a student body that is 38% Puerto Rican. Next in descending order of size is the African American mean percentage (i.e., 32.4), followed in turn by the European American (i.e., 17.1) and Asian/Pacific Islander American (i.e., 1.9) mean percentages. (The mean percentage for other ethnic/racial groups is 1.9; this variable is excluded from subsequent analyses.)

Table 1
Student Body's Ethnic/Racial Composition Variables: Means, Standard Deviations, Standard Errors of the Mean, and Skewness Values

Variable	M	SD	SEMean	Skewness
African American	32.4	28.7	3.08	0.58
Asian/Pacific Islander American	1.9	4.1	0.44	3.64
European American	17.1	26.8	2.90	1.91
Hispanic/Latino	46.5	28.8	3.14	0.16
Puerto Rican	(37.5)*	(25.9)	(3.05)	(0.37)
Other	[illegible]	[illegible]	[illegible]	[illegible]

Hispanic/Latino	(9.0)	(17.8)	(2.14)	(2.94)
Other ethnic/racial groups	1.9	6.4	0.69	4.97

Note. *N* = 84–87 schools. A school's measurement on a variable in this table is the percentage of the student body described by the variable. Percentages are within rounding error. [a]Estimated mean.

It also should be noted that the schools differ widely around these averages, as the standard deviations in Table 1 and the summary frequency distributions in Table 2 demonstrate. For example, Table 2 shows the following: About one fourth of the schools have a student body that is over 74% Hispanic/Latino, but at the other end of the distribution, another one fourth of the schools have a student body that is less than 25% Hispanic/Latino. About one third of the schools have a student body with an African American majority, but about one half of the schools have a student body that is less than 25% African American. About one tenth of the schools have a student body with a European American majority, but about three fourths of the schools have a student body that is less than 25% European American.

Table 2
Summary Frequency Distributions of Schools with respect to Student Body's Ethnic/Racial Composition

	African American[a]	**Asian/ Pacific Islander American[b]**	**European American[c]**	**Hispanic/ Latino[d]**
Percent of the school's student body	**Percent of schools**			
75% to 99%	10%	0%	7%	23%
50% to 74%	22	0	6	23
25% to 49%	17	1	9	29
0% to 24%	51	99	78	26

Note. *N* = 84–87 schools. The footnotes to this table describe the extremes of the tails of the distributions and other details. Percentages are within rounding error. [a]In 1% of the schools, the student body is 0.2% African American; in another 1% of the schools, the student body is 94.5% African American. In 30% of the schools, the majority (i.e., over 50%) of the student body is African

American. [b]In 48% of the schools, the number of Asian/Pacific Islander American students is zero; in 1% of the schools, the student body is 27% Asian/Pacific Islander American. In 99% of the schools, Asian/Pacific Islander Americans account for less than 15% of the student body. [c]In 7% of the schools, the number of European American students is zero; in 1% of the schools, the student body is 97.4% European American. In 12% of the schools, the majority of the student body is European American. [d]In 1% of the schools, the student body is 1.4% Hispanic/Latino; in another 1% of the schools, it is 98.7% Hispanic/Latino. In 43% of the schools, the majority of the student body is Hispanic/Latino.

2. What is the linguistic composition of the schools' student bodies?

The focal children attend schools in which, on average, monolingual native speakers of English constitute 58% of the student body; native speakers of Spanish, 36%; and native speakers of other languages, the remaining 5% (Table 3).

The correlation coefficients in Table 4 add to the evidence that schools tend to isolate students on the basis of both ethnicity/race and language.

The focal children attend schools in which, on average, students formally classified as limited-English-proficient (or English-language learners; LEP/ELL) constitute 18.5% of the student body (Table 3). This figure, when considered in relation to the mean percentages for the other linguistic- composition variables, shows that, on average in these schools, approximately 45% of the students who are not monolingual native speakers of English are formally classified as LEP/ELL.

Table 3
Student Body's Linguistic Composition Variables: Means, Standard Deviations, Standard Errors of the Mean, and Skewness

Variable	M	SD	SEMean	Skewness
Native speakers of Spanish	35.9	27.3	2.98	0.53
Monolingual native speakers of English	57.7	29.2	3.21	-0.32
Native speakers of other languages	5.2	12.5	1.38	4.88
Classified as LEP/ELL	18.5	13.3	1.44	0.74

Note. *N* = 82–86 schools. A school's measurement on a variable in this table is the percentage of the student body described by the variable. Percentages are within rounding error.

Table 4
Correlations among the Student Body's

Ethnic/Racial and Linguistic Composition Variables

Variable	2	3	4	5	6
Ethnic/racial composition					
1: African American	—	—	$-.25^{**}$	$.73^{***}$	$-.30^{**}$
2: European American	—	—	$-.21^{*}$	-.10	.05
3: Hispanic/Latino		—	$.89^{***}$	$-.28^{**}$	$.80^{***}$
Linguistic composition					
4: Native speakers of Spanish			—	$-.38^{***}$	$.74^{***}$
5: Monolingual native speakers of English				—	$-.32^{**}$
6: Classified as LEP/ELL					—

Note. N = 80–86 schools. The coefficients among the linguistic composition variables and the coefficients of variable 5 with variables 2 and 3 are Spearman rank-order correlations; the other coefficients in this table are Pearson product-moment correlations. The coefficients in this table are based on the variables measured in counts. $^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$ (1-tailed tests)

It also should be noted that the schools again vary widely around the mean percentages, as the standard deviations in Table 3 and the summary frequency distributions in Table 5 show. For example, native speakers of Spanish are the majority of the student body in about one third of the schools, but less than 25% of the student body in another one third of the schools. Similarly, monolingual native speakers of English constitute 75% or more of the student body in about one third of the schools, but less than 50% in another one third of the schools (Table 5).

Table 5
Summary Frequency Distributions of Schools on the Student Body's Linguistic Composition Variables

	Native speakers of Spanish[a]	Monolingual native speakers of English[b]	Native speakers of other languages[c]	Classified as LEP/ELL[d]
Percent of the school's student body	**Percent of schools**			
75% to				

99%	12%	36%	1%	0%
50% to 74%	19	27	1	4
25% to 49%	30	18	2	23
0% to 24%	39	19	95	73

Note. *N* = 82–86 schools. The footnotes to this table describe the extremes of the tails of the distributions and other details. Percentages are within rounding error. [a]In 1% of the schools, the student body is 0.2% native speaker of Spanish; in another 1% of the schools, the student body is 96.4% native speaker of Spanish. In 29% of the schools, the majority (i.e., over 50%) of the student body is native speaker of Spanish. [b]In 1% of the schools, the student body is 1.6% monolingual native speaker of English; in another 1% of the schools, it is 98.6% monolingual native speaker of English. In 58% of the schools, the majority of the student body is monolingual native speaker of English. [c]In 21% of the schools, there are zero native speakers of languages other than Spanish and English; in 1% of the schools, the student body is 88.7% native speakers of languages other than Spanish and English. [d]In 1% of the schools, there are zero students formally classified as LEP/ELL; in another 1% of the schools, 58% of the student body is formally classified as LEP/ELL.

3. What are the family socioeconomic characteristics of the schools' student bodies?

The schools have, on average, a student body composed largely of students who live in poverty and whose parents have very limited formal education, as Table 6 shows. Specifically, the mean percentages indicate that the schools have, on average, a student body characterized as follows: 42% of the students live in households in which the householder is unemployed; 45%, in households receiving public assistance (i.e., welfare); 60% of the students are eligible for fully subsidized lunch; and 68%, eligible for either fully or partly subsidized lunch. The mean for maternal education shows that the schools have, on average, a student body of which the average formal education level of the students' mothers or female guardians is below high school graduation (and below a General Education Diploma [GED]).

Table 6

Student Body's Family Socioeconomic Status Variables: Means, Standard Deviations, Standard Errors of the Mean, and Skewness Values

Variable	M	SD	SEMean	Skewness
Unemployment level	41.6	27.4	2.97	0.33
Public assistance dependence level	44.9	28.2	3.02	0.20
Economic status scale	1.43	0.60	0.06	1.41

Fully subsidized lunch eligibility level	59.8	25.9	2.83	-0.47
Partly subsidized lunch eligibility level	8.6	6.6	0.72	1.38
Subsidized lunch eligibility level (fully + partly)	68.4	26.8	2.94	-0.75
Maternal schooling scale	2.70	1.00	0.11	0.56

Note. N = 83–89 schools. A school's family *unemployment level* is the percentage of the student body living in households in which the householder is unemployed. *Public assistance dependence level* is the percentage of the student body from households receiving public assistance (i.e, welfare). The average family *economic status* of a school's student body is measured on a 5-point scale: 1 = low income; 2 = between middle and low income; 3 = middle income; 4 = between middle income and affluent; 5 = affluent. A school's *fully subsidized lunch eligibility level* is the percentage of the student body eligible for fully subsidized lunch. *Partly subsidized lunch eligibility level* is the percentage of the student body eligible for partly subsidized lunch. *Subsidized lunch eligibility level (fully + partly)* is the percentage of the student body eligible for fully subsidized lunch plus the percentage eligible for partly subsidized lunch. *Maternal schooling level* is the average level of formal education attained by the student body's mothers or female guardians, measured on a 9-point scale: 1 = six years of schooling or less; 2 = 7 to 9 years of schooling; 3 = 10 to 11 years; 4 = high school graduate or General Education Diploma (GED); 5 = post-high-school vocational or trade training; 6 = some college; 7 = college graduate; 8 = master's degree; 9 = doctor's degree.

Around each of these means is a wide range of differences among the schools, manifested in Tables 7 through 10. For example, in about two fifths of the schools, the student body is over 74% eligible for fully subsidized lunch, but at the other end of the distribution, in about one tenth of the schools, the student body is less than 25% thus eligible (Table 8). In one fifth of the schools, the student body is over 74% from homes with unemployed householders, but the student body is less than 25% from such homes in about one third of the schools (Table 7). In 8% of the schools, the student body's average maternal schooling level is less than a 7th-grade education, but in 17% of the schools it is high school graduation or a GED (Table 10).

Table 7
Summary Frequency Distributions of Schools on the Student Body's Family Unemployment Level and Public Assistance Dependence Level

	Unemployed householder[a]	**Household on public assistance**[b]
Percent of the school's student body	**Percent of schools**	
75% to 95%	20%	25%
50% to 74%	24	21
25% to 49%	22	23
1% to 24%	34	31

Note. $N = 85–87$ schools. The footnotes to this table describe the extremes of the tails of the distributions and other details. Percentages are within rounding error. [a]In 1% of the schools, the student body is 1% from households in which the householder is unemployed; in another 1% of the schools, the student body is 95% from such households. In 31% of the schools, the majority (i.e., over 50%) of the student body is from households in which the householder is unemployed. [b]In 2% of the schools, the student body is 1% from households receiving public assistance; in 1% of the schools, the student body is 95% from such households. In 37% of the schools, the majority of the student body is from households receiving public assistance.

Table 8
Summary Frequency Distributions of Schools on the Student Body's Subsidized Lunch Eligibility Variables

	Eligible for fully subsidized lunch[a]	**Eligible for partly subsidized lunch**[b]	**Eligible for subsidized lunch (fully + partly)**[c]
Percent of the school's student body	**Percent of schools**		
75% to 100%	39%	0%	52%
50% to 74%	26	0	25
25% to 49%	21	5	12
0% to 24%	13	95	11

Note. $N = 83–84$ schools. The footnotes to this table describe the extremes of the tails of the distributions and other details. Percentages are within rounding error. [a]In 1% of the schools, 2% of the student body is eligible for fully subsidized lunch; in another 1% of the schools, 99% of the student body is so.

In 65% of the schools, the majority (i.e., over 50%) of the student body is eligible for fully subsidized lunch. [b]In 1% of the schools, 0.1% of the student body is eligible for partly subsidized lunch; in another 1% of the schools, 31% of the student body is so. [c]In 1% of the schools, 3% of the student body is eligible for either fully or partly subsidized lunch; in 8% of the schools, 100% of the student body is so. In 77% of the schools, the majority of the student body is eligible for either fully or partly subsidized lunch.

Table 9
Frequency Distribution of Schools on the Student Body's Family Economic Status Scale

Student body's average family economic status	Percent of schools
Affluent	0%
Between middle income and affluent	1
Middle income	2
Between middle and low income	35
Low income	62
Note. *N* = 89 schools. Percentages are within rounding error.	

Table 10
Frequency Distribution of Schools on the Student Body's Maternal Schooling Scale

Student body's average maternal schooling level	Percent of schools	Cumulative percent
Doctor's degree	0%	0%
Master's degree	0	0
College graduate	0	0
Some college	1	1
Post-high school vocational or trade training	2	3
High school graduate or General Educ. Diploma (GED)	17	20
10 to 11 years	32	52
7 to 9 years	39	91
6 years or less	8	100
Note. *N* = 87 schools. Percentages are within rounding error.		

The intercorrelations among the student body's family socioeconomic variables show the expected pattern of consistency among measures of social, economic, and educational status (Table 11); these results add to the evidence supporting the data's construct validity.

Table 11
Intercorrelations among the Student Body's Family Variables

	2	3	4	5	6
1: Unemployment level	$.92^{***}$	$-.58^{***}$	$.75^{***}$	$.74^{***}$	$-.29^{**}$
2: Public assistance dependence level	--	$-.60^{***}$	$.80^{***}$	$.80^{***}$	$-.34^{***}$
3: Economic status scale		--	$-.52^{***}$	$-.52^{***}$	$.46^{***}$
4: Fully subsidized lunch eligibility level			--	$.98^{***}$	$-.36^{***}$
5: Subsidized lunch eligibility level (fully + partly)				--	$-.36^{***}$
6: Maternal schooling scale					--

Note. $N = 82–87$ schools. The coefficients of variable 3 with variables 1 and 2 are Spearman rank-order correlations; the other coefficients in this table are Pearson product-moment correlations. Variables 1, 2, 4, and 5 are measured in counts for the purpose of computing their intercorrelations; they are measured in percentages for the purpose of computing their correlations with variables 3 and 6. $^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$ (1-tailed tests)

4. In what types of neighborhoods are the schools located?

The schools are located mostly in highly urbanized areas—areas that are largely poor (Tables 12 and 13). Specifically, 60% of the schools are in the inner core of cities; 28%, in other urban parts of cities; 10%, in suburban neighborhoods; and 1%, in small towns. Forty-six percent (46%) of the schools are in low-income areas; 44%, in neighborhoods of a type characterized by a mix of low and middle income; 7%, in middle-income areas; and the remaining 3%, in neighborhoods comprising a mix of middle income and affluence (Table 13).

Table 12
School's Neighborhood Variables and School's Size and Crowdedness Variables: Means, Standard Deviations, Standard Errors of the Mean, and Skewness Values

	M	SD	SEMean	Skewness
School's neighborhood				
Urbanness scale	4.48	0.73	0.08	1.21
Economic status scale	1.67	0.75	0.08	1.11
School's size and crowdedness				
Enrollment size	677.2	295.8	31.4	0.39
Enrollment capacity	661.7[a]	265.8	29.2	0.38
Density-overcrowdedness level	15.5	205.2	22.5	0.44

Note. N = 88–89 schools for the school's neighborhood variables; N = 83–89 schools for the school's size and crowdedness variables. *Urbanness* is a 5-point scale: 1 = the school is in a rural area; 2 = small town (not suburban); 3 = suburban; 4 = urban part of a city other than its inner core; 5 = inner core of a city. The *economic status* of the neighborhood in which a school is located is measured on a 5-point scale: 1 = low income; 2 = mix of low and middle income; 3 = middle income; 4 = mix of middle income and affluent; 5 = affluent. A school's *enrollment size* is the total number of students enrolled in the school in late spring. *Enrollment capacity* is the number of students for which a school was built. A school's *density-overcrowdedness level* is measured by subtracting the enrollment capacity from the enrollment size; thus, a higher positive value signifies denser crowdedness than does a lower positive value. [a]Mean adjusted for missing data.

Table 13
Frequency Distributions of Schools on the Neighborhood Urbanness Scale and Neighborhood Economic Status Scale

Neighborhood urbanness scale		**Neighborhood economic status scale**	
School's location	**Percent of schools**	**School's location**	**Percent of schools**
Inner core of a city	60%	**Affluent area**	0%
Urban part of a city other than its inner core	28	**Mix of middle income and affluent**	3
Suburban	10	**Middle income**	7
Small town (not suburban)	1	**Mix of low and middle income**	44
Rural	0	**Low-income area**	46

Note. N = 88–89 schools. Percentages are within rounding error.

The correlations reported in Tables 14 and 15 show the following relationships: The more highly urbanized a school's neighborhood, the higher is the likelihood of the neighborhood's being poor. The lower a student body's average family economic status and parental schooling level, the higher is the likelihood of the school's being in an economically depressed and highly urbanized neighborhood.

Table 14
Correlations among the School's Neighborhood Variables and School's Size and Crowdedness Variables

	2	3	4	5	6
School's neighborhood					
1: Urbanness scale	-.63***	.36***	.34***	.10	.07
2: Economic status scale	—	-.25**	-.16	-.16	-.16
		School's size and crowdedness			
3: Enrollment size		—	.75***	.50***	.48***
4: Enrollment capacity			—	-.20*	-.04
5: Density-overcrowdedness level				—	.76***
6: Crowdedness dichotomy					—

Note. N = 83–89 schools. Pearson product-moment correlations. *Urbanness* is a 5-point scale: 1 = the school is in a rural area; 2 = small town (not suburban); 3 = suburban; 4 = urban part of a city other than its inner core; 5 = inner core of a city. The *economic status* of the neighborhood in which a school is located is measured on a 5-point scale: 1 = low income; 2 = mix of low and middle income; 3 = middle income; 4 = mix of middle income and affluent; 5 = affluent. A school's *enrollment size* is the total number of students enrolled in the school in late spring. *Enrollment capacity* is the number of students for which a school was built. A school's *density- overcrowdedness level* is measured by subtracting the enrollment capacity from the enrollment size; thus, a higher positive value signifies denser crowdedness than does a lower positive value. *Crowdedness dichotomy* is a dichotomous variable: 1 = the school is *not* crowded (i.e., density-overcrowdedness level is 0 or lower); 2 = the school is crowded (i.e., density- overcrowdedness level is greater than 0).
$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$ (1-tailed tests)

Table 15
Correlations of the Student Body's Family Variables with the School's Neighborhood Variables

	School's neighborhood variable	
Family variable	**Urbanness scale**	**Economic status scale**
Unemployment level	$.62^{***}$	$-.58^{***}$
Public assistance dependence level	$.53^{***}$	$-.60^{***}$
Economic status scale	$-.54^{***}$	$.74^{***}$
Fully subsidized lunch eligibility level	$.59^{***}$	$-.56^{***}$
Subsidized lunch eligibility level (fully + partly)	$.53^{***}$	$-.54^{***}$
Maternal schooling scale	$-.42^{***}$	$.42^{***}$

Note. N = 84–89 schools for the correlations of the school's neighborhood variables with the unemployment, public assistance, family economic status, and maternal schooling variables; N = 82–84 schools for the correlations of the neighborhood variables with the subsidized lunch variables. The coefficients of unemployment level and public assistance dependence level with the school's neighborhood variables are Spearman rank-order correlations; the other coefficients in this table are Pearson product-moment correlations. The unemployment, public assistance, and both subsidized lunch variables are measured in percentages. $^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$ (1-tailed tests)

5. What is the size of the schools? Are the school facilities crowded?

The schools have an average physical enrollment capacity for 662 students but enroll an average of 677 students (Tables 12 and 16). Forty-four percent (44%) of the schools enroll above capacity; that is, they enroll a higher number of students than the number for which the school was built (Table 17).

Table 16
Summary Frequency Distributions of Schools on Enrollment Size and Enrollment Capacity

	Enrollment size	Enrollment capacity
Number of students	Percent of schools	
1,200 to 1,400	4%	5%
1,000 to 1,199	17	8
800 to 999	14	23
600 to 799	17	24
400 to 599	27	23
200 to 399	20	16

86 to 199	1	1
Note. *N* = 83–89 schools. Percentages are within rounding error.		

Table 17
Summary Frequency Distribution of Schools on Density-Overcrowdedness Level

School's density-overcrowdedness level	**Percent of schools**	**Cumulative percent**
600 to 680	2%	2%
400 to 599	0	2
200 to 399	17	19
1 to 199	25	44
0	5	49
-1 to -199	40	89
-200 to -399	10	99
-400 to -515	1	100

Note. *N* = 83 schools. A school's *density-overcrowdedness level* is measured by subtracting the enrollment capacity from the enrollment size; thus, a higher positive value signifies denser crowdedness than does a lower positive value. Percentages are within rounding error.

There are, however, wide differences among the schools on each of these variables, as Tables 16 and 17 show. For example, 13% of the schools have a capacity for as many as 1,000 to 1,400 students, but 17% of the schools, for fewer than 400. Twenty-one percent (21%) of the schools enroll as many as 1,000 to 1,400 students, but another 21%, fewer than 400 (Table 16). Nineteen percent (19%) of the schools enroll 200 or more students above capacity, but 51% of the schools enroll below capacity (Table 17).

The correlations in Tables 14 and 18 show the following: The larger a school, the higher is the likelihood of its being located in a highly urbanized, economically impoverished area. Also, the larger a school, the lower is its student body's average parental schooling level, and the higher is its student body's family unemployment rate.

Table 18
Correlations of the Student Body's Family Characteristics with the School's Size and Crowdedness

	School's size and crowdedness			
Family variable	**Enrollment size**	**Enrollment capacity**	**Density-overcrowd level**	**Crowdedness dichotomy**
Unemployment level	$.18^{*}$	$.20^{*}$	.06	.02
Public assistance dependence level	.12	.15	.02	.03
Economic status scale	-.16	-.13	-.06	-.12
Fully subsidized lunch eligibility level	.09	.06	.06	.06
Subsidized lunch eligibility level (fully + partly)	.10	.02	.12	.11
Maternal schooling scale	$-.24^{**}$	$-.27^{**}$	.01	-.04

Note. $N = 77\text{–}89$ schools. Pearson product-moment correlations. The unemployment, public assistance, and both subsidized lunch variables are measured in percentages. $^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$ (1-tailed tests)

6. Correlates of the student body's ethnic/racial composition:

6.1. What are the relationships of the student body's ethnic/racial composition to the student body's family socioeconomic characteristics?

The relative concentration of Hispanics/Latinos in the student body correlates positively with the student body's family unemployment level, public assistance dependence level, and subsidized lunch eligibility level and, congruent with these relationships, negatively with the student body's family economic status scale and maternal schooling scale. This pattern of correlations is largely similar to the pattern of relationships between the relative concentration of African American students and these measures of the student body's socioeconomic characteristics. These correlations are in a direction *opposite* to that of the correlations between the relative concentration of European American students and these measures of the student body's socioeconomic characteristics. In short, these analysis results, reported in Table 19, signify the following:

The higher a school's concentration of Hispanic/Latino pupils, the lower is the student body's average family socioeconomic status and parental schooling level. Similarly, the higher the concentration of African American pupils, the lower is the student body's average family socioeconomic status. In contrast, the higher the concentration of European American students, the more affluent and the more highly educated, on average, are the student body's families.

Table 19
Correlations of the Student Body's Ethnic/Racial Composition with the Student Body's Family, School's Neighborhood, and School's Size and Crowdedness Characteristics

	African American	European American	Hispanic/Latino
Family[a]			
Unemployment level	$.47^{***}$	$-.41^{***}$	$.52^{***}$
Public assistance dependence level	$.47^{***}$	$-.38^{***}$	$.55^{***}$
Economic status scale	$-.21^{*}$	$.58^{***}$	$-.38^{***}$
Fully subsidized lunch eligibility level	$.32^{**}$	$-.30^{**}$	$.61^{***}$
Subsidized lunch eligibility level (fully + partly)	$.31^{**}$	$-.24^{*}$	$.64^{***}$
Maternal schooling scale	.04	$.39^{***}$	$-.43^{***}$
School's neighborhood[b]			
Urbanness scale	$.25^{**}$	$-.69^{***}$	$.46^{***}$
Economic status scale	$-.22^{*}$	$.54^{***}$	$-.34^{***}$
School's size and crowdedness[c]			
Enrollment size	-.11	-.16	$.25^{**}$
Enrollment capacity	.00	$-.18^{*}$	.08
Density-overcrowd level	$-.24^{*}$	-.02	$.30^{**}$
Crowdedness dichotomy	$-.19^{*}$	-.10	$.28^{**}$

[a]$N = 79\text{–}87$ schools for the coefficients involving the family variables. The coefficients of the African American variable with the family variables, and the coefficients of the ethnic/racial composition variables with the family economic status scale and the maternal schooling scale are Pearson product-moment correlations; the coefficients of the ethnic/racial composition variables with the other family variables are Spearman rank-order correlations. The unemployment, public assistance, and both subsidized lunch variables are measured in counts for the purpose of computing their correlations in this table; likewise, the ethnic/racial composition variables are measured in counts for the purpose of computing their correlations with the unemployment, public assistance, and both subsidized lunch variables. The ethnic/racial composition variables are measured in percentages for the purpose of computing their correlations with the other variables in this table. [b]$N = 83\text{–}87$ schools for the coefficients involving the school's neighborhood variables. The coefficients of the ethnic/racial composition variables with the school's neighborhood variables are Pearson product-moment correlations. [c]$N = 78\text{–}87$ schools for the coefficients involving the school's size and crowdedness variables. The coefficients of the ethnic/racial composition variables with the crowdedness dichotomy are Pearson product-moment correlations; the coefficients of the ethnic/racial composition variables with the other school size and crowdedness variables are Spearman rank-order correlations. $^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$ (1-tailed tests)

6.2. What are the relationships of the student body's ethnic/racial composition to the characteristics of the school's neighborhood?

The correlations in Table 19 show the following: The higher the concentration of Hispanic/Latino students in a school, the higher is the likelihood of the school's location being an economically depressed and highly urbanized area. An association similar to this occurs between the relative concentration of African American students and these school neighborhood characteristics. In contrast, the higher the concentration of European American students in a school, the *lower* is the likelihood of the school's being located in a poor or highly urbanized neighborhood.

6.3. **Is the student body's ethnic/racial composition related to school size and crowdedness?**

There is little or no relationship between ethnic/racial composition and school size. On the other hand, the student body's percentage of Hispanics/Latinos correlates positively and significantly with the school crowdedness dichotomy (Table 19). These analyses thus show that schools with higher proportions of Hispanic/Latino students are more likely to be crowded (i.e., more likely to enroll in excess of the number of pupils for which the school was built) than schools with lower proportions of this ethnic/racial group.

7. Correlates of the student body's linguistic composition:

7.1. What are the relationships of the student body's linguistic composition to the student body's family socioeconomic characteristics?

The student body's relative concentration of native speakers of Spanish correlates positively with the student body's family unemployment level, public assistance dependence level, and subsidized lunch eligibility level and, consistent with these associations, negatively with the student body's family economic status scale and maternal schooling scale. These correlations are similar to those between the student body's relative concentration of LEP/ELL students and these measures of the student body's socioeconomic characteristics. In contrast, the student body's relative concentration of monolingual native speakers of English correlates *positively* with the student body's family economic status scale and maternal schooling scale. These results, presented in Table 20, signify the following:

The higher a school's concentration of pupils who are native speakers of Spanish, the lower is the student body's average family socioeconomic status and parental schooling level. Similarly, the higher a school's concentration of LEP/ELL pupils, the lower is the student body's average family socioeconomic status and parental schooling level. In contradistinction, the higher a school's concentration of pupils who are monolingual native speakers of English, the *higher* is the student body's average family economic status and parental schooling level.

Table 20
Correlations of the Student Body's Linguistic Composition with the Student Body's Family, School's Neighborhood, and School's Size and Crowdedness Characteristics

	Native speakers of Spanish	**Monolingual native speakers of English**	**Classified as LEP/ELL**
Family[a]			
Unemployment level	$.54^{***}$	.12	$.38^{***}$
Public assistance dependence level	$.57^{***}$	.10	$.40^{***}$
Economic status scale	$-.35^{***}$	$.25^{**}$	$-.25^{**}$
Fully subsidized lunch eligibility level	$.62^{***}$	.13	$.53^{***}$
Subsidized lunch eligibility level (fully + partly)	$.65^{***}$	.10	$.54^{***}$
Maternal schooling scale	$-.35^{***}$	$.33^{***}$	$-.25^{**}$
School's neighborhood[b]			
Urbanness scale	$.38^{***}$	$-.34^{***}$	$.42^{***}$
Economic status scale	$-.32^{***}$	$.24^{*}$	$-.28^{**}$
School's size and crowdedness[c]			

School's size and crowdedness			
Enrollment size	.18*	-.28**	.12
Enrollment capacity	.07	-.08	.04
Density-overcrowd level	.25**	-.37***	.19*
Crowdedness dichotomy	.24*	-.33**	.08

[a]*N* = 79–86 schools for the coefficients involving the family variables. The coefficients of the linguistic composition variables with the family economic status scale and the maternal schooling scale are Pearson product-moment correlations; the coefficients of the linguistic composition variables with the other family variables are Spearman rank-order correlations. The unemployment, public assistance, and both subsidized lunch variables are measured in counts for the purpose of computing their correlations in this table; likewise, the linguistic composition variables are measured in counts for the purpose of computing their correlations with the unemployment, public assistance, and both subsidized lunch variables. The linguistic composition variables are measured in percentages for the purpose of computing their correlations with the other variables in this table. [b]*N* = 82–86 schools for the coefficients involving the school's neighborhood variables. The coefficients of the linguistic composition variables with the school's neighborhood variables are Pearson product- moment correlations. [c]*N* = 79–86 schools for the coefficients of the linguistic composition variables with the school's size and crowdedness variables. The coefficients of the linguistic composition variables with the crowdedness dichotomy are Pearson product-moment correlations; the coefficients of the linguistic composition variables with the other school size and crowdedness variables are Spearman rank-order correlations. $^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$ (1-tailed tests)

7.2. What are the relationships of the student body's linguistic composition to the characteristics of the school's neighborhood?

Table 20 shows the following relationships: The higher a school's concentration of students who are native speakers of Spanish, the higher is the likelihood of the school's location being a low-income, inner-city area. Similarly, the higher a school's concentration of LEP/ELL students, the higher is the likelihood of its location being a poor, highly urbanized area. In contrast, the higher a school's concentration of students who are monolingual native speakers of English, the higher is the likelihood that its location is in the more affluent and less urbanized neighborhoods.

7.3. Is the student body's linguistic composition related to school size and crowdedness?

Table 20 shows that the school crowdedness dichotomy correlates positively with the student body's percentage of native speakers of Spanish, but negatively with the student body's percentage of monolingual native speakers of English. Enrollment capacity is not related to the student body's linguistic composition. These results demonstrate the following relationships: The larger a school's proportion of pupils who are native speakers of Spanish, the higher is the school's likelihood of being crowded. In contrast, the larger a school's proportion of pupils who are monolingual native speakers of English, the lower is its likelihood of being crowded.

Discussion

In this century, few issues in North America have aroused more intense and bitter controversy, or caused more renting and sustained conflict, than those surrounding ethnic/racial integration generally and school desegregation in particular (see, e.g., Lukas, 1986; Woodward, 1966). At present, more than a century after *Plessy v. Ferguson* and almost half a century after *Brown v. Board of Education*, the fundamental concerns remain unresolved in practice; indeed, they have grown in complexity. In 1896, in the *Plessy* decision, the U.S. Supreme Court codified racial segregation, making it the law of the land. In 1954, in the *Brown* decision, the Court reversed the *Plessy* decision. Current trends, however, point to a *de facto* return to widespread segregated schooling, as the present study shows.

In recent years, the public debate concerning education reform in the United States has given relatively little attention to certain critical attributes of the ecology of schooling, particularly to attributes that bear on the isolation of students by ethnicity/race, language, and family socioeconomic characteristics. These attributes of schooling—and their interrelationships—were examined in the present study, focusing specifically on the schools that children who migrate from Puerto Rico to New Jersey (i.e., focal children) attend in the United States during the first two years following their arrival in this country.

This study shows that there is considerable ethnic/racial segregation of students in many of the schools attended by focal children. Hispanics/Latinos are the majority of the student body in 43% of the schools. European Americans are the majority of the student body in only 12% of the schools. This study further shows that there is considerable isolation by language. Native speakers of Spanish are the majority of the student body in nearly one third of the schools.

Economic impoverishment and low parental education are also salient attributes of the student body in many of the schools. In 65% of the schools, the majority of the student body is eligible for fully subsidized lunch. In addition, many of the schools are located in highly urbanized and economically depressed areas. Nearly two thirds of the schools are in the inner core of cities; most of the remaining third, in other urban parts of cities. Almost one half are in low-income areas.

As used here in reference to the present study's findings, the term *school segregation*, or *school isolation*, does *not* necessarily imply that the school boards or other public school officials caused the ethnic/racial, linguistic, or socioeconomic segregation of students observed in the present study. Regardless of the causes, however, the observed patterns of segregation do not bode well. Insofar as a school does not provide adequate occasions for interethnic interactions, it deprives students of the opportunity to develop the sociocultural knowledge, shared understandings, and behavior patterns that they will need as adults in order to function harmoniously and productively in ethnically heterogeneous settings (Laosa, 1999)—a serious problem for a society as increasingly diverse as ours. Other potential consequences of the observed patterns of ethnic/racial and linguistic isolation are discussed in subsequent sections of this article.

The present findings gain in significance in the light of previous research suggesting an influence of the student body's socioeconomic status on scholastic achievement (Kennedy et al., 1986, chap. 2; Myers, 1985; Orland, 1994; Puma et al., 1993; U.S. Department of Education, 1993b, 1996, 1997). One may further hypothesize that the ecology of schools can affect not only a child's academic achievement but also his or her long-term social development. For instance, a neighborhood with a high unemployment rate will likely provide limited exposure to successfully employed role models (Brooks-Gunn, Denner, & Klebanov, 1995; Laosa, 1999; Wilson, 1995). Children in such schools are largely cut off from a range of options and opportunities commonly available in middle-class schools.

Based on the available research evidence, a U.S. Department of Education (1993b) report concluded that "teachers in high-poverty schools face special challenges that often undermine their effectiveness" (p. 31). Although studies clearly confirm a relationship between student body poverty and academic achievement, the evidence is weaker concerning the mechanisms, or processes, that may explain this relationship (see, e.g., Barton et al., 1991; Taylor & Piché, 1991; and U.S. Department of Education, 1993b, 1996, 1997, for reviews of research). The data collected in the larger investigation of which the present study is a part will permit analyses to illuminate these processes.

A large size and crowdedness are additional attributes of many schools attended by focal children. The schools attended by the focal children enroll an average of 677 pupils—a much larger figure than the estimated average number of pupils per public elementary school for the United States nationwide, for New Jersey and New York statewide, and for Puerto Rico island-wide; respectively they are 458, 419, 582, and 298 (U.S. Department of Education, 1993a, Table 96). Moreover, 44% of the focal children's schools enroll in excess of the number of pupils for which they were built. These findings must be considered in light of the potential effects of school size and crowdedness on the focal children's academic performance and socioemotional adjustment—an issue for future research. Also needed is research concerning the effects on the focal children of the dramatic size difference between the schools they attend in this country and those in Puerto Rico. Additional issues for future research are considered later.

Separation and Inequality

The student body's ethnic/racial composition and linguistic composition were found to correlate with the student body's socioeconomic characteristics, with school crowdedness, and with the school neighborhood's characteristics. The larger a school's proportion of pupils who are Hispanic/Latino or native speakers of Spanish, the higher is the school's concentration of pupils from economically impoverished and poorly educated parents, and the higher its likelihood of being crowded and of being located in an economically depressed and highly urbanized area. Similarly, the larger a school's proportion of African American pupils, the higher is its concentration of pupils from low-income families and the higher its likelihood of being in a poor inner-city area. In contrast, the larger a school's proportion of European American pupils, the *lower* is its concentration of pupils from economically impoverished and poorly educated parents, and the *lower* its likelihood of being in an economically depressed and highly urbanized area.

The correlational analyses thus clearly show that separate is *not* equal. School segregation by ethnicity/race is closely associated with school segregation by poverty and by parental education. Similarly, school segregation by language is closely associated with school segregation by poverty and by parental education. Furthermore, ethnic/racial segregation and linguistic segregation are associated with crowded schools.

A focal child in a school with a relatively high concentration of pupils who are Hispanic/Latino or native speakers of Spanish is likely in a school with a high concentration of pupils from economically impoverished and poorly educated families, a crowded school located in a poor inner-city area. In contrast, a focal child in a school with a relatively high proportion of European American pupils is likely in a school with relatively *few* students from economically impoverished or poorly educated families, a school that is *not* located in an economically depressed or highly urbanized area.

The present findings raise crucial questions concerning equality of educational opportunity, fairness, and social justice— concerns that urgently need the attention of educators, parents, and policy makers. Equal educational opportunity is the fundamental American answer to social and economic inequality, but school segregation by ethnicity/race or language does in effect concentrate poverty and low academic achievement in schools that are not equal—a historical and contemporary fact (e.g., Barton et al., 1991; Bremner, Barnard, Hareven, & Mennel, 1970, 1971, 1974; Forehand, Ragosta, & Rock, 1976; Kennedy et al., 1986; Laosa, 1984; Orfield, 1993; Orland, 1994; Puma et al., 1993; Taylor & Piché, 1991; U.S. Department of Education, 1993b, 1996, 1997). Such schools are often vulnerable to becoming overwhelmed with problems of economically impoverished and poorly educated families isolated in neighborhoods lacking many of the opportunities typically available in other schools. The challenging task of providing access for these children to appropriate and effective schooling so that every student can have a fair chance of becoming a full participant in American society demands high priority (Cárdenas, 1995, 1996; Donato et al., 1991; Network of Regional Desegregation Assistance Centers, 1989; Orfield, 1993; Orfield et al., 1996; Orfield & Yun, 1999).

Differences Among the Schools

It is also important to note that substantial differences among the focal children's schools occur on almost all the variables. The schools differ widely in student body ethnic/racial composition. For example, in about one fourth of the schools, Hispanics/Latinos constitute between 75% and 99% of the student body; yet at the other end of the distribution, in another one fourth of the schools, they constitute less than 25% of the student body. In about one tenth of the schools, European Americans constitute 50% to 98% of the student body, although in about three quarters of the schools they are less than 25% of the student body.

Similarly, the schools differ widely in linguistic composition. For instance, in about one third of the schools, native speakers of Spanish are the majority of the student body, but in about two fifths of the schools they are less than 25% of the student body.

The schools also differ widely in student body socioeconomic characteristics, school size, and density-overcrowdedness. In addition, although to a lesser extent, the schools differ with regard to quality of location.

Needed Research

From the perspective of scientific inquiry, the observed differences among the focal children's schools constitute a series of naturally occurring experiments, raising a compelling question: Will these differences among the schools explain, or statistically predict, individual differences in focal children's learning and adaptation? The present findings point to specific hypotheses in need of systematic research, as next steps in the larger longitudinal investigation of which this study is a part. For example, concerning the potential influence of the observed ecological attributes of schools on particular dimensions of child outcome, the following hypotheses focus on language development:

The *second-language* ***motivation*** *hypothesis* predicts that the strength of the motivation to acquire a second language will vary as a function of the need to communicate through that language. If this hypothesis is correct, then the larger a school's concentration of pupils who are native speakers of Spanish, the weaker will be a focal child's need to use English to communicate with peers, hence the lower the child's motivation to learn English, and hence the slower the child's English-language development rate.

The *second-language* ***exposure*** *hypothesis* predicts that the rate of learning a second language will depend on the exposure to that language (i.e., on the frequency, or probability, of opportunities to hear and use the language in functional situations). This hypothesis predicts a relatively slow rate of English-language development in the schools with relatively small proportions of pupils who are monolingual speakers of English. Thus, both hypotheses make the same prediction, namely, a negative relationship between the student body's proportion of native speakers of Spanish and focal children's English-language development rate.

On the other side of the coin is the *native-language loss hypothesis*. According to it, second-language learners will, to the extent that they have limited opportunity to use their native language actively, lose native-language skills (Laosa, 1999). If this hypothesis is accurate, then the smaller a school's proportion of Spanish-speaking students, the fewer will be the focal child's opportunities to use Spanish, and hence the faster the rate of Spanish- language loss.

Especially for the focal population, development of *both* languages is vitally important: English-language development is, of course, critically important for children's academic achievement and psychosocial adaptation in the United States. Because of the special relationship between the two countries, many focal children return to Puerto Rico—establishing a "circular migration" pattern—where they must compete (in school and eventually in the workplace) through the Spanish language. Thus, especially for them, continued Spanish-language development is as critically important as English-language acquisition.

Language development and academic achievement are not the only child outcomes that the school ecology may influence. Psychosocial/affective outcomes may also be influenced. Various hypotheses bear on this point. For instance, according to the *intercultural stress hypothesis*, the cultural "distance" (i.e., the degree of difference) between ecological settings bears on psychosocial adaptation (Laosa, 1999). This hypothesis predicts that the wider the difference between the child's primary culture/language and the school context, the more exacting and hence the more stressful and anxiety-producing will be the school experience. In turn, these high levels of psychological distress will raise the probability of behavioral/emotional problems. If this hypothesis is valid, then focal children in schools with relatively few Hispanic/Latino pupils who are native speakers of Spanish will show a higher prevalence of symptoms of behavioral/affective maladjustment than will the focal children in schools with larger proportions of such pupils.

In short, for focal children, the consequences of relatively intense levels of ethnolinguistic segregation (i.e., high concentrations of Hispanic/Latino, native-Spanish-speaking pupils) may include relatively slow rates of English- language development, but little or no loss of Spanish, and a relatively high probability of healthy behavioral/emotional adjustment. These hypotheses thus illustrate some of the difficult dilemmas that one must confront when addressing the question, What is best for a focal child? These and other hypotheses can be tested using the longitudinal data from the larger investigation of which this study is a part—an investigation uniquely designed to permit this important and urgently needed scientific research.

School Segregation Policies and Judicial Trends in the United States

According to some historians (e.g., Woodward, 1966), the doctrines of Anglo-Saxon superiority by which some intellectuals and politicians justified and rationalized U.S. imperialism in the Caribbean, Latin America, and the Pacific did not differ in essentials from the race theories espoused by those who sought to justify White supremacy over African Americans. In 1896, two years before the United States acquired Puerto Rico, the U.S. Supreme Court's ruling in the case of *Plessy v. Ferguson* affirmed a vision of a rigidly segregated society. Homer Plessy—of mixed African and European ancestry—had taken an East Louisiana Railway train car seat reserved for Whites; (Note 16) as a consequence, he was jailed for violating a segregation statute that forbade members of either race to occupy accommodations set aside for the other—with the exception of "nurses attending the children of the other race" (as quoted in Kunen, 1996, p. 40). Segregation statutes, or "Jim Crow" laws, constituted a strict code that, as Woodward (1966) noted, "lent the sanction of law to a racial ostracism that extended to churches and schools, to housing and jobs, to eating and drinking. Whether by law or by custom, that ostracism extended to virtually all forms of public transportation, to sports and recreations, to hospitals, orphanages, prisons, and asylums, and ultimately to funeral homes, morgues, and cemeteries" (p. 7). In a nearly unanimous decision on *Plessy*, the Supreme Court declared that laws mandating "equal but separate" treatment of the races "do not necessarily imply the inferiority of either race," and cited the widely accepted propriety of separate schools for White and "colored" children. In lone dissent, Justice John Harlan remarked, "The thin disguise of 'equal' accommodations . . . will not mislead anyone, nor atone for the wrong this day done" (as quoted in Kunen, 1996, p. 40).

From 1896 to 1954 northern and southern state policies and practices confirmed the prediction that Justice Harlan had made in his dissenting opinion in *Plessy*: that the Court's decision would place "in a condition of legal inferiority a large body of American citizens" (as quoted in F. C. Jones, 1981, p. 72). The thin disguise to which he referred endured for a half century until African American plaintiffs in a series of court cases challenged the constitutionality of school segregation (Orfield et al., 1996; Woodward, 1966). The plaintiffs in these cases were attacking not only inequality, but segregation itself (Woodward, 1966). These cases culminated in the 1954 Supreme Court's landmark decision in *Oliver Brown et al. v. Board of Education of Topeka, Kansas*, (Note 17) which reversed a constitutional trend begun long before *Plessy*. The new Chief Justice, Earl Warren, delivered the Court's unanimous opinion in favor of the African American plaintiffs: "We conclude," said the Chief Justice, "that in the field of public education, the doctrine of `separate but equal' has no place. Separate educational facilities are inherently unequal." The plaintiffs had therefore been "deprived of the equal protection of the laws guaranteed by the Fourteenth Amendment" of the U.S. Constitution; consequently, intentional segregation in public schools was unconstitutional (as quoted in Woodward, 1966, p. 147). By thus ruling that *de jure* segregation was unlawful, the *Brown* decision reversed the *Plessy* decision, which rested on the principle that there could be "separate-but- equal" treatment of people (Laosa, 1984; Sitkoff, 1993; Woodward, 1966).

Central to the promise inherent in the *Brown* decision is the belief that ethnic/racial segregation in public education has a detrimental effect on children and "may affect their hearts and minds in a way unlikely ever to be undone" (as quoted in Woodward, 1966, p. 147)—not because ethnically/racially segregated institutions are inherently inferior but due to continuing structural inequities directly attributable to ethnic/racial prejudice and discrimination (E. R. Jones, 1996).

In the first decade after *Brown* very little desegregation occurred in the South (Rist, 1979). There was open defiance and massive resistance against attempts to implement the *Brown* mandate (Motley, 1995; Sitkoff, 1993; Woodward, 1966). The federal government and the federal district courts in the South did little to pressure the states or the school districts to comply with the constitutional requirements of the *Brown* decision (Orfield et al., 1996; van Geel, 1982, p. 980; Zashin, 1978). Moreover, segregation in the North remained virtually untouched until the 1970s. According to Orfield et al. (1996, p. 8), "Most Northern districts even refused to provide racial data that could be used to measure segregation." For nearly two decades following *Brown*, the Supreme Court denied hearings to school desegregation cases from the North (Note 18) (Orfield et al., 1996), a historical fact illustrating that the legal meaning of desegregation has evolved (see, e.g., Kirp, 1977; Landsberg, 1995; Orfield, 1978; Orfield et al., 1996; van Geel, 1982).

Although the Supreme Court's decision in *Brown* greatly encouraged many Hispanics/Latinos, it did not offer definitive guidance on how to combat discrimination against them (González, 1982; Laosa, 1984). Various issues have arisen in desegregation litigation involving this ethnic/racial group, all hinging on the identifiability of the group and of its members (Levin, Castaneda, & von Euler, 1977; Orfield, 1978; Orfield et al., 1996; Roos, 1977). A central question the courts have asked in judging whether the isolation of Hispanic/Latino students violates the equal protection clause of the Fourteenth Amendment is whether Hispanics/Latinos constitute a group (i.e., a "class") that should be legally treated in the same manner as African Americans (Levin et al., 1977; Roos, 1977). In other words, Are Hispanics/Latinos a group such that discrimination against them violates the equal protection clause? Schools, courts, and policy makers were uncertain how to categorize Hispanics/Latinos for the purposes of civil rights (González, 1982).

In the mid- 1960s momentous changes began to occur: Martin Luther King, Jr., and his organization marched in the early 1960s, and in so doing raised the moral conscience of the nation (Laosa, 1984; Oates, 1982; van Geel, 1982). The administrations of presidents John F. Kennedy and Lyndon B. Johnson provided executive leadership in the battle for civil rights. In 1964 the U.S. Congress passed the Civil

Rights Act, which required cutting off federal funds to school districts and other institutions that discriminate: Title VI of the Act states, "No person in the United States shall, on the ground of race, color, or national origin, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance" (78 Stat. 252 [1964]; 42 U.S.C. 2000d [1965]).

An important key to questions of how to combat discrimination against Hispanic/Latino students appeared in the Civil Rights Act of 1964. This law and the authorization it vested on federal agencies to enforce it "by issuing rules, regulations, or orders of general applicability" established a legal basis to regulate matters pertaining to national origin discrimination in addition to race (Civil Rights Act of 1964, as quoted in González, 1982, p. II-3). This law gave federal education officials responsibilities for working with the courts to enforce the *Brown* decision and subsequent decisions requiring racial desegregation. To this end, the then Office of Education (OE) of the U.S. Department of Health, Education, and Welfare (HEW) developed guidelines to ensure compliance with Title VI. Aiding OE's efforts, Congress passed the Elementary and Secondary Education Act of 1965, which substantially increased the amount of federal assistance to public education, thereby making fund cutoffs a more serious threat (Laosa, 1984; Zashin, 1978).

The Supreme Court, too, provided strong leadership on desegregation during that period. For example, in 1968, the Court declared that discrimination must be "eliminated root and branch" (*Green v. County School Board of New Kent County*, as quoted in Orfield et al., 1996, p. xxii). In 1971, the Court held in *Swann v. Charlotte-Mecklenburg Board of Education* and in *North Carolina State Board of Education v. Swann* that the federal courts could order busing to desegregate schools (Orfield, 1978; Orfield et al., 1996; Zirkel, Richardson, & Goldberg, 1995).

Despite this country's long history of persistent school segregation and other forms of discrimination against Hispanic/Latino students (see, e.g., Carter & Segura, 1979; Donato, Menchaca, & Valencia, 1991; González, 1982; Laosa, 1984; U.S. Commission on Civil Rights, 1971, 1972; Weinberg, 1977), the task of proving to the courts that these discriminatory practices are *de jure* rather than *de facto* was frequently more difficult for this ethnic/racial group than for African Americans. (Note 19) In cases involving discrimination against African Americans in the South, previous state statutes or constitutional provisions requiring segregation of this group had usually existed, and they were widely known and understood and could be readily documented (Laosa, 1984; Orfield, 1978). In order to establish a case of unlawful segregation, therefore, African American plaintiffs have needed merely to show the continued presence of school segregation in school systems formerly segregated by law (Levin et al., 1977; van Geel, 1982). In contrast, Hispanic/Latino plaintiffs have frequently been hindered by a lack of systematic documentation concerning the magnitude of educational exclusion of their group and by unclear understandings of the policies underlying the group's disenfranchisement (González, 1982).

In the absence of a statutory history of *de jure* segregation, Hispanic/Latino plaintiffs in segregation cases have been required to show that they are segregated and that the segregation is attributable to intentional action by school officials or other state authorities. In other words, proving to the courts that the isolation of Hispanic/Latino students constitutes a violation of the equal protection clause has required a showing of *de jure* segregation attributable not to statute but instead to the action of school officials (Levin et al., 1977; Roos, 1977). For example, in *United States v. Texas Education Agency* (1972, as cited in Levin et al., 1977) the circuit court found intentional segregative action by the school district, particularly in the choice of school sites, construction of schools, drawing of attendance zones, and student assignment and transfer policies. The court thus found *de jure* segregation of Hispanic/Latino students despite the absence of a previous statute requiring segregation of this ethnic/racial group, and stated that discrimination in this case was "no different from any other school desegregation case" (as quoted in Levin et al., 1977, p. 76). (Note 20)

The U.S. Supreme Court did not begin to try to untangle the problem of school segregation as it relates to Hispanics/Latinos until 1973, when it tried the case of *Keyes v. School District No. 1* (Denver, Colorado). In *Keyes* the Supreme Court recognized the problem but did not solve it entirely, seemingly saying that at least some Hispanics/Latinos, in some regions, under some conditions, should be recognized as a distinct class:

> There is also much evidence that in the Southwest Hispanos and Negroes have a great many things in common. . . . Though of different origins, Negroes and Hispanos in Denver suffer identical discrimination in treatment when compared with the treatment afforded Anglo students. In that circumstance, we think petitioners are entitled to have schools with a combined predominance of Negroes and Hispanos included in the category of "segregated" schools. (*Keyes*, 413 U.S. 189 [1973], as quoted in González, 1982, p. II-7)

In multi-ethnic areas, this recognition has often meant that the degree of segregation in a school depends on the ratio of European American students to the combined number of identified "minority" students in that school (Levin et al., 1977; Roos, 1977). Issues left unresolved by the Supreme Court's ruling in *Keyes* were articulated by Orfield (1978, pp. 203-204):

> The [*Keyes*] decision mentions conditions prevailing in the Southwest. It is unclear whether the same rights extend to Mexican- Americans in cities outside the

> Southwest. Would evidence that social conditions had changed in a part of the Southwest remove this special constitutional protection for Mexican-American children? Conditions in the region vary greatly on matters ranging from residential segregation to intermarriage, socioeconomic mobility to educational achievement. It is not clear what factors would determine how a particular Hispanic group in a given part of the country should be treated for desegregation purposes.

Although a narrow reading could indeed limit applicability to Mexican Americans/Chicanos in the Southwest, in applying *Keyes* the courts have often "interpreted this aspect of the holding expansively, neither restricting application of the term Hispanic to Chicanos in the Southwest nor requiring a showing of `identical discrimination'" (Teitelbaum & Hiller, 1977, p. 165). Subsequent to *Keyes*, courts in school desegregation cases have typically treated children from other Hispanic/Latino groups—and from certain other ethnic/racial groups as well—as "minority" students (Teitelbaum & Hiller, 1977, p. 165). For example, federal judges in New York and Boston decided that desegregation could be extended to Hispanic/Latino groups that were primarily Puerto Rican (Orfield, 1978, p. 204; Teitelbaum & Hiller, 1977, p. 165).

More broadly, *Keyes* is also significant because, as the Supreme Court's first case on desegregation in the "North," it expanded desegregation requirements to the North and West (Orfield et al., 1996). (Note 21) Before 1970, legal developments had not affected racial segregation patterns outside the South because such patterns had usually been characterized as de facto. In the 1970s, however, the courts were finding—as the Supreme Court did in the *Keyes* case in Denver—that much northern urban segregation was *de jure* segregation based not on statute but instead on specific acts or policies of school boards and other school officials (Brown, 1995; Orfield, 1978).

In the early 1970s, public protests intensified over the potential expansion of school desegregation and over forced transportation (i.e., busing) of students as a means to desegregate. Accordingly, the leadership that the executive and legislative branches of government were providing in desegregation efforts waned. Moreover, by this time, as a consequence of demographic alterations in the ethnic/racial composition of the U.S. population and shifts in residential patterns, many Northern urban school districts, which seldom extend beyond city limits, lacked sufficient numbers of European American children to desegregate (Kunen, 1996; Orfield, 1978). By the time of President Richard Nixon's second term of office, significant progress toward school desegregation had virtually stopped (Orfield et al., 1996; Orfield, 1978; Orfield & Monfort, 1992).

In 1974, the Supreme Court began issuing a series of decisions limiting *Brown*'s reach. For example, in *Milliken v. Bradley* [1974] the

Supreme Court erected serious barriers to interdistrict, city-suburban desegregation plans; such plans have aimed to desegregate racially isolated schools that are located in urban areas by drawing students from the surrounding suburban districts. In this Detroit metropolitan case, the Supreme Court prohibited such plans unless plaintiffs could demonstrate that the suburbs or the state took actions that contributed to segregation in the city. Because obtaining such legal proof is often difficult, *Milliken* seriously limits access to the option of drawing students from largely European American suburbs in order to desegregate urban districts that enroll high concentrations of students of color (Orfield et al., 1996). That unconstitutional segregation existed in Detroit was not questioned in this case; in question was the constitutionality of the court- ordered desegregation plan's extending to outlying districts with no history of segregative action on the part of their school boards or local governments (Zirkel et al., 1995). Throughout the country, large numbers of students of color are segregated in urban areas; hence, insofar as *Milliken* puts suburban schools out of reach of these students, it practically ensures their isolation in the cities (Orfield et al., 1997; Orfield & Monfort, 1992; van Geel, 1982).

During the 1980s, the executive branch of the federal government worked actively against mandatory school desegregation; and Congress accepted a proposal from President Ronald Reagan's administration to slash the budget for federal desegregation assistance programs (Orfield et al., 1996). In recent years, neither branch has made a significant school desegregation initiative.

In *Milliken v. Bradley II* [1977] the Supreme Court, facing the challenge of providing a remedy for the Detroit schools, where *Milliken I* had made long-term integration practically impossible, had ruled that a court could order a state to pay for educational programs to repair the harms caused by segregation (Orfield et al., 1996; Zirkel et al., 1995). More recently, however, in *Missouri v. Jenkins* [1995], the Supreme Court ruled that the court-ordered programs designed to improve the quality of education in predominantly poor, predominantly non-White schools in order to make them educationally more equal to other schools, and to increase the attractiveness of schools in order to accomplish desegregation through voluntary choices, should be temporary, and that school districts need not show any actual correction of the educational harms of segregation before such programs can be discontinued (Orfield et al., 1996, 1997). Analyzing this court decision, Orfield and his colleagues (1996, p. xv) concluded that the Supreme Court by allowing, as it did in this case, for the dismantling of the special educational programming that the district had established as a remedy for students in segregated schools, may have signaled that in the future the Court may not even support enforcement of the "separate but equal" doctrine that *Brown* overturned. That is, it seems reasonable to conclude from the apparent underlying philosophy in the Supreme Court's rulings in *Jenkins* and in two other recent cases (i.e., *Board of Education of Oklahoma City v. Dowell* in 1991 and *Freeman v. Pitts* in 1992) that, in issues of school desegregation, the U.S. Supreme Court as presently constituted is pursuing the twin goals of minimizing judicial involvement in education and quickly restoring authority to local and state government, "whatever the consequences" (Orfield et al., 1996, p. 3).

In sum, the urgent focus of public opinion on civil rights lasted only two years, from 1963 to 1965. Vigorous and effective enforcement of school desegregation by the executive branch of the federal government began in 1965 and lasted four years (González, 1982; Laosa, 1984; Orfield et al., 1996). The Supreme Court continued to provide strong leadership on desegregation for four more years, in a series of sweeping decisions from 1969 to 1973—decisions that launched busing as a remedy, extended desegregation requirements from the South to northern cities, established the right of Hispanic/Latino children to desegregated schools, and declared that it was no longer permissible to delay implementing the Court's mandate to desegregate (González, 1982; Orfield, 1978; Orfield & Monfort, 1992; Rist, 1979; Zirkel et al., 1995). Congressional leadership on civil rights weakened after 1965 as public opinion changed. Efforts toward school desegregation then waned on the part of the three branches of government. Political and legal forces have converged in recent years to effect movement in a direction opposite to that of efforts to desegregate public education (Orfield et al., 1996, 1997; Orfield & Yun, 1999).

School Segregation Trends in the United States

A clear correspondence can be seen, on the one hand, between the foregoing chronology of events pertaining to efforts to desegregate American schools and, on the other, the annual national statistics on the segregation of African American students: During the 1964-1972 period of active enforcement in the southern and border states, a major decline occurred in the segregation of those regions' African American students. The South changed from almost total segregation in 1963 to become the most desegregated region of the country by 1970 (Orfield & Monfort, 1988; Rist, 1979). (Note 22) In the early 1970s the trend toward increased desegregation of African American students virtually stopped. Then, in 1988, a drift toward increased segregation of African American students began (Orfield, 1993; Orfield et al., 1996, 1997; Orfield & Yun, 1999). The corresponding national statistics on the segregation of Hispanic/Latino students show, however, a strikingly different trend, as noted below.

Studies by Orfield and his colleagues and by other researchers show a steady trend in the United States toward increased school segregation of Hispanic/Latino children. This trend is evident since national data on the subject were first collected, in the 1960s. Indeed, since 1980 Hispanics/Latinos have been more likely than African Americans to attend predominantly minority schools. (Note 23) Specifically, nationwide in the 1968- 69 academic year, 77% of African American students and 55% of Hispanic/Latino students attended predominantly minority schools; in 1972-73 these figures were 64% and 57%; by 1980-81 they had switched to 63% and 68%. In 1996-97, 69% of African American students and 75% of Hispanic/Latino students attended predominantly minority schools (Orfield, 1993; Orfield et al., 1997; Orfield & Yun, 1999). A similar trend can be observed in other measures of segregation, namely, the percentage of children of each ethnic/racial group in schools with a 90% to 100% minority enrollment (Orfield, 1993; Orfield et al., 1997; Orfield & Yun, 1999; U.S. Department of Education, 1995), and the weighted average percentage of European American students in the schools attended by children of a particular ethnic/racial group (Orfield, 1993; Orfield et al., 1997; Orfield & Yun, 1999).

Needed: Public Awareness, Policies, and Leadership

Some advocates of bilingual education for Hispanic/Latino children have sometimes objected to efforts to desegregate students from this ethnolinguistic group, fearing that such desegregation may weaken support for the bilingual/bicultural education programs that many of these children need. Other advocates and experts on the subject have argued that there is *no* inherent conflict between bilingual/multicultural education and desegregation, that under certain conditions both can be effectively realized—indeed, and that with sufficient will and effort, the aims of both can be achieved synergistically to produce educationally successful, integrated communities. There is an urgent need to inform parents, educators, and policy makers of the reality, the issues, the potential consequences, and the as-yet- unanswered questions about the existing segregation of ethnolinguistic minority children in our nation's schools.

Heretofore, solutions to the problems of school segregation have been sought almost exclusively through the courts. Certainly, the most significant advances toward desegregation of African American students have been achieved with the considerable help of judicial decisions. At present, however, the problems of school segregation are even more complex and difficult than those of the past. There is also growing evidence that these problems affect multiple ethnic/racial and linguistic groups (perhaps in different ways), including children who migrate from Puerto Rico, as this study shows. Some observers have questioned whether the courts (particularly as they are presently constituted), and the adversarial system on which the judicial structure rests, are still the most effective and appropriate means possible for policy formation in an area as complex as school segregation (cf. Cárdenas, 1995; Fischer, 1982). Be that as it may, it is now painfully evident that desegregation does not guarantee integration, nor ensure full equality of educational opportunity (Brown, 1995; Cárdenas, 1995; Laosa, 1984, 1999; Teitelbaum & Hiller, 1977).

It seems clear, considering the statistical trends and the history of school desegregation efforts, that significant advances in solving problems of school segregation cannot in the foreseeable future be achieved through the courts *alone*. Urgently needed are creative, informed efforts toward the formulation of comprehensive solutions, and concerted leadership to implement them effectively.

Notes

1. For editorial simplicity, the term country is used here as if Puerto Rico and the United States were two distinct countries. Following this usage, the terms United States (U.S.) and American(s) are used exclusively in reference to the 50 states (and the District of Columbia) of the United States and the people therein. Similarly, the term Hispanic/Latino is used exclusively to refer to the Hispanic/Latino population of the 50 states (and the District of Columbia). The present usage does not imply any view regarding Puerto Rico's sociopolitical status, which at present is neither that of an independent nation nor that of a state of the United States. Of the 50 states, New Jersey has the highest Puerto Rican population density and the second-largest proportion of the total Puerto Rican population that resides stateside (Pérez & Martínez, 1993; U.S. Bureau of the Census, 1992, 1993).

2. Giving rise to these developments were several significant ideological, economic, and political currents in the United States: As the end of the nineteenth century approached, there were changes in thought about the nation's mission and its destiny. The nation had become a world power because of its prodigious economic growth (Link, 1992; Morison, 1972). After the disappearance of the "American frontier," the conviction grew that the country needed to find new outlets for an ever increasing population and agricultural and industrial production. Advocates of sea power argued that "future national security and greatness" depended upon a large navy supported by bases throughout the world (Link, 1992, p. 248). Social Darwinists advanced the view that the world is a jungle, with international rivalries inevitable, and that only a strong nation could survive (Link, 1992; Morison, 1972). Added to these arguments were those of idealists and religious leaders who believed that Americans had a duty to "take up the White man's burden" and to carry their assertedly superior culture "to the backward peoples of the world" (Link, 1992, p. 248; Morison, 1972; Woodward, 1966). It was against this background that the Spanish-American War of 1898 propelled the United States along the road to war and empire (Lewis, 1963; Link, 1992; Morison, 1972)—a war that, although brief and relatively bloodless, had far- reaching and long-lasting political and diplomatic consequences. These overseas incursions brought under the nation's jurisdiction some eight million people of color, "a varied assortment of inferior races," as the Nation described them, "which, of course, could not be allowed to vote" (1898, as quoted in Woodward, 1966, p. 72).

3. More specifically, schools with at least one third- or fourth-grade class (or the equivalent for ungraded programs). This study focuses on public and not private schools because a previous study (Laosa, 1998) showed that of the total population of elementary-school transfers-in from Puerto Rico to New Jersey, only a tiny proportion are transfers-in to non-public schools.

4. Below are the annual distributions of children transferring in from Puerto Rico to the third and fourth grades (or the equivalent for ungraded programs) in the sample of New Jersey schools. To avoid inflating these counts, if a child transferred in from Puerto Rico more than once during the course of the investigation, the child was counted only once.

Number of children	Number of schools	
	Year 1	Year 2
0	169	177
1	27	21
2	16	8
3	9	8
4	5	9
5	4	4
6	5	4
7	3	3
8	0	3
9	2	1
10	0	2
11	0	1
12	0	0
13	0	0
14	1	0

5. The data describe the school at the time that focal children attended it; if the school had focal children more than one academic year, then the analyses selected the data corresponding to the first academic year that the school had focal children.

6. Counts rather than percentages were used in computing this variable's correlations with certain other variables; see footnote 15.

7. Counts rather than percentages were used in computing this variable's correlations with certain other variables; see footnote 15.

8. Counts rather than percentages were used in computing this variable's correlations with certain other variables; see footnote 15.

9. Consistent with the usage adopted by the U.S. Bureau of the Census, the term householder (rather than head of household) is used in the presentation of data that had previously been presented with the designation head (e.g., U.S. Bureau of the Census, 1994b, p. A-2).

10. Counts rather than percentages were used in computing this variable's correlations with certain other variables; see footnote 15.

11. Counts rather than percentages were used in computing this variable's correlations with certain other variables; see footnote 15.

12. Counts rather than percentages were used in computing this variable's correlations with certain other variables; see footnote 15.

13. Counts rather than percentages were used in computing this variable's correlations with certain other variables; see footnote 15.

14. Two matrices of correlation coefficients were computed: a matrix of Pearson product-moment correlations and a matrix of Spearman rank-order correlations; depending on the shape of the observed frequency distributions on a given pair of variables, either one type of coefficient or the other is reported; the two coefficients are very similar or practically identical to each other for the vast majority of the pairs of variables. Variables with distributions too skewed to yield meaningful coefficients were excluded from the correlation matrices.

15. To avoid the spurious correlation that may occur between variables that share in common the same variable denominator (McNemar, 1969, pp. 180-182), whenever two variables shared in common the same variable denominator, the correlation

between them was computed using counts rather than percentages. The Appendix presents the descriptive statistics based on counts for these variables.

16. In the United States, persons of mixed European and African ancestry are generally considered Black/African American (i.e., "non-White"). This system of racial classification differs from the predominant conceptions of race and of racial identification in Puerto Rico; for a discussion of these conceptions see Rodríguez (1991).

17. Four separate cases from Kansas, South Carolina, Virginia, and Delaware were consolidated and decided in the 1954 case of Brown v. Board of Education. In each case, African Americans sought admission to the public schools of their community on a nonsegregated basis. Kansas, by state law, permitted but did not require segregated schools. The other three states had state constitutional and statutory provisions that required the segregation of Blacks and Whites in public schools (Zirkel, Richardson, & Goldberg, 1995).

18. The nature of racial segregation in the North differed from that in the South: Typically in the South, school segregation was required by state constitutional or statutory provisions.

19. The term "*de jure* segregation" generally refers to segregation that has had the sanction of law; that is, segregation directly intended by law or otherwise issuing from an official racial classification. The term comprehends situations in which the activities of school authorities have had a racially discriminatory impact contributing to the establishment or continuation of school segregation. The term "*de facto* segregation" is limited to what is "inadvertent and without the assistance or collusion of school authorities" and not caused by state action (Black, Nolan, Nolan-Haley, Connolly, Hicks, & Alibrandi, 1990, pp. 416, 425). State action refers to action by the government, including action by a public school system or its agents (Zirkel et al., 1995, p. 208).

20. Similarly, in Cisneros v. Corpus Christi Independent School District (1970, Texas), the circuit court had found *de jure* segregation to exist, noting that the

> de jure nature of the existing pattern of segregation within the Corpus Christi Independent School District has as its basis state action of a non-statutory variety—that is, the school board's active pursuit of policies that not only do nothing to counteract the effect of existing patterns of

> residential segregation in view of viable alternatives of significant integrative value, but, in fact, increase and exacerbate the district's racial and ethnic imbalance. There has been a history of official school board acts which have had such a segregative effect. (Cisneros, 1970, as quoted in Levin et al., 1977, p. 76)

Thus, once the necessary intentional segregative actions were found, coupled with a high concentration of Hispanic/Latino students in some schools, a *prima facie* case of unlawful segregation was established (Levin et al., 1977).

Cisneros is the first circuit court case to hold that Hispanics/Latinos must be considered an identifiable minority group for purposes of desegregation; that is to say, that the principles enunciated in Brown v. Board of Education apply to Hispanics/Latinos as well as to African Americans. This decision prevented school officials in Corpus Christi from claiming that they had desegregated a school by placing in it only African American and Hispanic/Latino (i.e., Mexican American) students (González, 1982; Levin et al., 1977).

21. Keyes is the first Supreme Court opinion addressing de jure segregation in a city (Denver, Colorado) located in a state where at the time of Brown v. Board of Education the public schools were not segregated pursuant to state statutory authority (Brown, 1995, p. 650). Many of Denver's public schools were segregated, although the city's school system had never been operated under a state constitutional provision or law that mandated or permitted school segregation (Zirkel et al., 1995, p. 113).

22. Significantly, prior to 1964 no systematic data on the implementation of Brown were collected. The general consensus among those who studied this period is that fewer than 1% of all African American students in the eleven southern states attended desegregated schools (i.e., schools that White/European American students also attended; Rist, 1979, p. 4). In the same academic year (1964-65) of the passage of the Civil Rights Act, the first private efforts at collecting desegregation data on these states began. The findings from those efforts suggest that 2% of all African American students in these states were in desegregated schools. In 1965-66 the federal government began to collect data; that year, 7% of the South's African American students were in desegregated schools (Rist, 1979, p. 4). Then the pace of desegregation in the South quickened: The first national statistics on school desegregation became available with the 1968-69 academic year. That year 23% of African American students nationwide were in majority-White schools, in contrast with 18% in the South alone. Within two years the

shift was dramatic as the South had 39% of its African American students in majority-White schools, compared with 28% in the northern and western states (Orfield, 1978, pp. 56-57; Orfield & Monfort, 1992, p. 13; Rist, 1979, p. 4).

23. A predominantly minority school is one in which more than half of the school's combined enrollment is African American, American Indian/Native American, Asian/Pacific Islander American, or Hispanic/Latino (Orfield, 1993, p. 5).

References

Barton, P. E., Coley, R. J., & Goertz, M. E. (1991). *The state of inequality* (Policy Information Report). Princeton, NJ: Educational Testing Service.

Black, H. C., Nolan, J. R., Nolan-Haley, J. M., Connolly, M. J., Hicks, S. C., & Alibrandi, M. N. (1990). *Black's law dictionary* (6th ed.). Saint Paul, MN: West Publishing.

Bremner, R. H., Barnard, J., Hareven, T. K., & Mennel, R. M. (Eds.). (1970). *Children and youth in America: A documentary history: Vol 1. 1600-1865.* Cambridge, MA: Harvard University Press.

Bremner, R. H., Barnard, J., Hareven, T. K., & Mennel, R. M. (Eds.). (1971). *Children and youth in America: A documentary history: Vol 2. 1866-1932.* Cambridge, MA: Harvard University Press.

Bremner, R. H., Barnard, J., Hareven, T. K., & Mennel, R. M. (Eds.). (1974). *Children and youth in America: A documentary history: Vol 3. 1933-1973.* Cambridge, MA: Harvard University Press.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design.* Cambridge, MA: Harvard University Press.

Bronfenbrenner, U. (1995). Developmental ecology through space and time: A future perspective. In P. Moen, G. H. Elder, Jr., & K. Lüscher (Eds.), *Examining lives in context: Perspectives on the ecology of human development* (pp. 619-647). Washington, DC: American Psychological Association.

Brooks-Gunn, J., Denner, J., & Klebanov, P. (1995). Families and neighborhoods as contexts for education. In E. Flaxman & A. H. Passow (Eds.), *Changing populations, changing schools. Ninety-fourth yearbook of the National Society for the Study of Education* (Part 2, pp. 233-252). Chicago: University of Chicago Press.

Brown, K. (1995). Revisiting the Supreme Court's opinion in *Brown v. Board of Education* from a multiculturalist perspective. *Teachers College Record, 96,* 644-653.

Cárdenas, J. A. (1995). *Multicultural education: A generation of advocacy.* Needham Heights, MA: Simon & Schuster Custom Publishing.

Cárdenas, J. A. (1996). Ending the crisis in the K- 12 system. In L. I. Rendón & R. O. Hope and associates (Eds.), *Educating a new majority: Transforming America's educational system for diversity* (pp. 51-70). San Francisco: Jossey-Bass.

Carr, R. (1984). *Puerto Rico: A colonial experiment.* New York: Random House.

Carter, T. P., & Segura, R. D. (1979). *Mexican Americans in school: A decade of change.* New York: College Entrance Examination Board.

Cobb, C. D., & Glass, G. V (1999). Ethnic segregation in Arizona charter schools. *Education Policy Analysis Archives* [On-line journal], *7*(1). Available: http://epaa.asu.edu/epaa/v7n1/

Donato, R., Menchaca, M., & Valencia, R. R. (1991). Segregation, desegregation, and integration of Chicano students: Problems and prospects. In R. R. Valencia (Ed.), *Chicano school failure and success: Research and policy agendas for the 1990s* (pp. 27-63). London: Falmer.

Fischer, L. (1982). The courts and educational policy. In A. Lieberman & M. W. McLaughlin (Eds.), *Policy making in education. Eighty-first yearbook of the National Society for the Study of Education* (pp. 56-79). Chicago: University of Chicago Press.

Forehand, G. A., Ragosta, M., & Rock, D. A. (1976). *Conditions and processes of effective school desegregation* (Report No. PR-76-23. Final report prepared for the Office of Education, U.S. Department of Health, Education, and Welfare). Princeton, NJ: Educational Testing Service.

González, J. M. (1982). *Hispanics, bilingual education and desegregation: A review of major issues and policy directions.* Unpublished report to the U.S. Commission on Civil Rights.

Jones, E. R. (1996). Foreword. In G. Orfield, S. E. Eaton, & the Harvard Project on School Desegregation, *Dismantling desegregation: The quiet reversal of Brown v. Board of Education* (pp. vii-ix). New York: The New Press/Norton.

Jones, F. C. (1981, Spring). External crosscurrents and internal diversity: An assessment of black progress, 1960- 1980. *Daedalus*, 71-101.

Kennedy, M. M., Jung, R. K., & Orland, M. E. (1986). *Poverty, achievement, and the distribution of compensatory education services.* Washington, DC: U.S. Government Printing Office.

Kirp, D. L. (1977). Law, politics, and equal educational opportunity: The limits of judicial involvement. *Harvard Educational Review, 47*, 117-137.

Kunen, J. S. (1996, April 29). The end of integration. *Time*, pp. 39-45.

Landsberg, B. K. (1995). The federal government and the promise of *Brown*. *Teachers College Record, 96*, 627-636.

Laosa, L. M. (1984). Social policies toward children of diverse ethnic, racial, and language groups in the United States. In H. W. Stevenson & A. E. Siegel (Eds.), *Child development research and social policy* (pp. 1- 109). Chicago: University of Chicago Press.

Laosa, L. M. (1990). Psychosocial stress, coping, and development of Hispanic immigrant children. In F. C. Serafica, A. I. Schwebel, R. K. Russell, P. D. Isaac, & L. B. Myers (Eds.), *Mental health of ethnic minorities* (pp. 38-65). New York: Praeger.

Laosa, L. M. (1997). Research perspectives on constructs of change: Intercultural migration and developmental transitions. In A. Booth, A. Crouter, & N. Landale (Eds.), *Immigration and the family: Research and policy on U.S. immigrants* (pp. 133-148). Mahwah, NJ: Erlbaum.

Laosa, L. M. (1998). *Child migration from Puerto Rico to public and private schools in the United States: Sampling a difficult-to-reach population* (Research Rep. No. 98-24). Princeton, NJ: Educational Testing Service.

Laosa, L. M. (1999). Intercultural transitions in human development and education. *Journal of Applied Developmental Psychology, 20*(3), 355-406.

Laosa, L. M. (n.d.). *Psychosocial stress and Hispanic immigrant children's coping and adaptation to their role as students: Puerto Rican migration. Progress report.* Princeton, NJ: Educational Testing Service.

Laosa, L. M., & Henderson, R. W. (1991). Cognitive socialization and competence: The academic development of Chicanos. In R. R.

Valencia (Ed.), *Chicano school failure and success: Research and policy agendas for the 1990s* (pp. 164-199). New York: Falmer.

Levin, B., Castaneda, S., & von Euler, M. (1977). Legal issues related to school desegregation and the educational concerns of the Hispanic community. Desegregation and education concerns of the Hispanic community. Conference report (pp. 75-89). Washington, DC: National Institute of Education, U.S. Department of Health, Education, and Welfare.

Lewis, G. K. (1963). *Puerto Rico: Freedom and power in the Caribbean.* New York: Monthly Review Press.

Link, A. S. (1992). Imperialism, the progressive era, and the rise to world power, 1896-1920. *The new encyclopaedia Britannica: Macropaedia* (15th ed., Vol. 29, pp. 248-253). Chicago: Encyclopaedia Britannica.

Lukas, J. A. (1986). *Common ground: A turbulent decade in the lives of three American families.* New York: Random House.

Mathews, T. G., & Tata, R. J. (1992). Puerto Rico. *The new encyclopaedia Britannica: Macropaedia* (15th ed., Vol. 29, pp. 755-761). Chicago: Encyclopaedia Britannica.

McNemar, Q. (1969). *Psychological statistics* (4th ed.). New York: Wiley.

Minuchin, P. P., & Shapiro, E. K. (1983). The school as a context for social development. In E. M. Hetherington (Ed.), *Handbook of child psychology: Vol. 4. Socialization, personality, and social development* (pp. 197-274). New York: Wiley.

Moore, J., & Pachon, H. (1985). *Hispanics in the United States.* Englewood Cliffs, NJ: Prentice-Hall.

Morison, S. E. (1972). *The Oxford history of the American people: Vol. 3. 1869-1963.* New York: New American Library.

Motley, C. B. (1995). The legacy of *Brown v. Board of Education. Teachers College Record, 96,* 637-643.

Myers, D. E. (1985). The relationship between school poverty concentration and students' reading and math achievement learning. In M. M. Kennedy, R. K. Jung, & M. E. Orland (1986), *Poverty, achievement, and the distribution of compensatory education services* (Appendix D, pp. D-16 to D-60). Washington, DC: U.S. Government Printing Office.

Network of Regional Desegregation Assistance Centers. (1989). Resegregation of public schools: The third generation. Portland, OR: Author/Northwest Regional Educational Laboratory.

Oates, S. B. (1982). *Let the trumpet sound: The life of Martin Luther King, Jr.* New York: New American Library.

Orfield, G. (1978). *Must we bus? Segregated schools and national policy.* Washington, DC: Brookings Institution.

Orfield, G. (1993). The growth of segregation in American schools: Changing patterns of separation and poverty since 1968. Alexandria, VA: Council of Urban Boards of Education, National School Boards Association.

Orfield, G., Bachmeier, M., James, D. R., & Eitle, T. (1997). Deepening segregation in American public schools. Cambridge, MA: Harvard Project on School Desegregation, Harvard University.

Orfield, G., Eaton, S. E., & the Harvard Project on School Desegregation. (1996). *Dismantling desegregation: The quiet reversal of Brown v. Board of Education.* New York: The New Press/Norton.

Orfield, G., & Monfort, F. (1988). Racial change and desegregation in large school districts: Trends through the 1986-1987 school year. Alexandria, VA: Council of Urban Boards of Education, National School Boards Association.

Orfield, G., & Monfort, F. (1992). Status of school desegregation: The next generation. Alexandria, VA: Council of Urban Boards of Education, National School Boards Association.

Orfield, G., & Yun, J. T. (1999). Resegregation in American schools. Cambridge, MA: Civil Rights Project, Harvard University.

Orland, M. E. (1994). Demographics of disadvantage: Intensity of childhood poverty and its relationship to educational achievement. In J. I. Goodlad & P. Keating (Eds.), *Access to knowledge: The continuing agenda for our nation's schools* (Rev. ed., pp. 43- 58). New York: College Entrance Examination Board.

Pérez, S. M., & Martínez, D. (1993). State of Hispanic America 1993: Toward a Latino anti-poverty agenda. Washington, DC: National Council of La Raza.

Puma, M., Jones, C. C., Rock, D., & Fernandez, R. (1993). Prospects: The congressionally mandated study of educational growth and opportunity (Interim report prepared for the U.S. Department of

Education, Planning and Evaluation Service). Bethesda, MD: Abt Associates.

Rist, R. C. (1979). Introduction. In R. C. Rist (Ed.), *Desegregated schools: Appraisals of an American experiment* (pp. 1-11). New York: Academic Press.

Rodríguez, C. E. (1991). *Puerto Ricans born in the U.S.A.* Boulder, CO: Westview Press.

Roos, P. (1977). Issues in desegregation remedial order for Hispanos. Desegregation and education concerns of the Hispanic community. Conference report (pp. 29-35). Washington, DC: National Institute of Education, U.S. Department of Health, Education, and Welfare.

Rutter, M., Maughan, B., Mortimore, P., & Ouston, J. (1979). *Fifteen thousand hours: Secondary schools and their effects on children.* Cambridge, MA: Harvard University Press.

Sitkoff, H. (1993). *The struggle for Black equality: 1954-1992* (Rev. ed.). New York: Hill and Wang.

Southern Education Foundation, Panel on Educational Opportunity and Postsecondary Desegregation. (1995). Redeeming the American promise: Report of the Panel on Educational Opportunity and Postsecondary Desegregation. Atlanta, GA: Author.

Taylor, W. L., & Piché, D. M. (1991). Shortchanging children: The impact of fiscal inequity on the education of students at risk (Report prepared for the Committee on Education and Labor of the U.S. House of Representatives; Serial No. 102-O). Washington, DC: U.S. Government Printing Office.

Teitelbaum, H., & Hiller, R. J. (1977). Bilingual education: The legal mandate. *Harvard Educational Review, 47,* 138-170.

U.S. Bureau of the Census. (1992). 1990 Census of population. General population characteristics: United States (1990 CP-1-1). Washington, DC: U.S. Government Printing Office.

U.S. Bureau of the Census. (1993). 1990 Census of population and housing. Population and housing unit counts: United States (1990 CPH-2-1). Washington, DC: U.S. Government Printing Office.

U.S. Bureau of the Census. (1994a). Educational attainment in the United States: March 1993 and 1992 (Current Population Reports, Series P20-476; by R. Kominski & A. Adams). Washington, DC: U.S. Government Printing Office.

U.S. Bureau of the Census. (1994b). The Hispanic population in the United States: March 1993 (Current Population Reports, Series P20-475; by P. A. Montgomery). Washington, DC: U.S. Government Printing Office.

U.S. Bureau of the Census. (1996). Statistical abstract of the United States: 1996 (116th ed.). Washington, DC: U.S. Government Printing Office.

U.S. Commission on Civil Rights. (1971). Ethnic isolation of Mexican Americans in the public schools of the Southwest (Report 1 of the Mexican American Education Study). Washington, DC: U.S. Government Printing Office.

U.S. Commission on Civil Rights. (1972). The excluded student: Educational practices affecting Mexican Americans in the Southwest (Report 3 of the Mexican American Education Study). Washington, DC: U.S. Government Printing Office.

U.S. Commission on Civil Rights. (1976). Puerto Ricans in the continental United States: An uncertain future. Washington, DC: Author.

U.S. Department of Education. (1993a). Digest of education statistics, 1993 (NCES 93-292). Washington, DC: National Center for Education Statistics, U.S. Government Printing Office.

U.S. Department of Education. (1993b). Reinventing Chapter 1: The current Chapter 1 program and new directions (Rep. of the National Assessment of Chapter 1 Independent Review Panel). Washington, DC: U.S. Department of Education, Office of Policy and Planning.

U.S. Department of Education. (1995). Findings from The Condition of Education 1995: No. 4. The educational progress of Hispanic students (NCES 95-767, by T. M. Smith). Washington, DC: National Center for Education Statistics, U.S. Government Printing Office.

U.S. Department of Education. (1996). Urban schools: The challenge of location and poverty. Executive summary (NCES 96-864, by L. Lippman, S. Burns, & E. McArthur, with contributions by R. Burton, T. M. Smith, & P. Kaufman). Washington, DC: National Center for Education Statistics, U.S. Government Printing Office.

U.S. Department of Education. (1997). Findings from The Condition of Education 1997: No. 10. The social context of education (NCES 97-981, by B. A. Young & T. M. Smith). Washington, DC: National Center for Education Statistics, U.S. Government Printing Office.

U.S. General Accounting Office. (1992). Remedial education: Modifying Chapter 1 formula would target more funds to those most in need (GAO/HRD-92-16). Washington, DC: Author.

van Geel, T. (1982). Judicial decisions. In H. E. Mitzel, J. H. Best, & W. Rabinowitz (Eds.), *Encyclopedia of educational research* (5th ed., Vol. 2, pp. 974- 982). New York: The Free Press.

Wagenheim, K. (1970). *Puerto Rico: A profile.* New York: Praeger.

Weinberg, M. (1977). *A chance to learn: A history of race and education in the United States.* New York: Cambridge University Press.

Wilson, W. J. (1995). Jobless ghettos and the social outcome of youngsters. In P. Moen, G. H. Elder, Jr., & K. Lüscher (Eds.), *Examining lives in context: Perspectives on the ecology of human development* (pp. 527-543). Washington, DC: American Psychological Association.

Woodward, C. V. (1966). *The strange career of Jim Crow* (2nd Rev. ed.). Oxford, England: Oxford University Press.

Zashin, E. (1978, Winter). The progress of Black Americans in civil rights: The past two decades assessed. *Daedalus,* 239-262.

Zirkel, P. A., Richardson, S. N., & Goldberg, S. S. (1995). *A digest of Supreme Court decisions affecting education* (3rd ed.). Bloomington, IN: Phi Delta Kappa Educational Foundation.

About the Author

Luis M. Laosa
Principal Research Scientist
Educational Testing Service
Turnbull Hall, 8-R
Rosedale Road
Princeton, New Jersey 08541

Email: llaosa@ets.org
Phone: (609) 734-5524

Luis M. Laosa has conducted extensive research in varied Hispanic/Latino communities (Chicano/Mexican American, Puerto Rican, Cuban American) throughout the United States and in Mexico and South America. His current studies include a large-scale longitudinal project focusing on child migration, supported in part by the William T. Grant Foundation and the Spencer Foundation. He is the author of numerous scientific and scholarly publications and is often sought as scientific and technical advisor to government agencies, universities, research centers, and philanthropic foundations. He is a fellow of the American Psychological Society and of the American Psychological Association (in the divisions of developmental psychology, educational psychology, general psychology, and ethnic-minority psychology). Other honors include a Martin Luther King, Jr./César Chávez/Rosa Parks Visiting Professorship at the University of Michigan, receipt of the Educational Testing Service's Senior Scientist Award, and induction into the Phi Kappa Phi Honor Society. Dr. Laosa has served on the editorial boards of *Review of Educational Research, Child Development, Developmental Psychology, Journal of Educational Psychology, Psychological Bulletin*, the *Journal of Applied Developmental Psychology, Early Education and Development*, and *Journal of School Psychology*. He received his Ph.D. (1971) from the University of Texas at Austin (specializing in cross-cultural psychology, personality/social development, educational psychology, measurement, and research methodology); completed a postdoctoral residency in clinical and community psychology at the University of Texas Medical School, Health Sciences Center, at San Antonio; and received his certification in school psychology and his professional certification and license in general psychology. He was chief school psychologist in a large school district and has served on the clinical faculty of the Department of Psychiatry of the University of Texas Medical School and on the faculty of the Graduate School of Education of the University of California, Los Angeles. He has been on the research staff of Educational Testing Service in Princeton, New Jersey, since 1976, where at present he holds the post of principal research scientist.

Acknowledgment

The research presented here was made possible in part by a research grant from the William T. Grant Foundation to the author. The data presented, the statements made, and the views expressed are solely the responsibility of the author. This study is part of the author's large-scale longitudinal research project focusing on children who migrate to the United States from Puerto Rico

Appendix

Descriptive Statistics for Variables Measured in Counts: Means, Standard Deviations, Standard Errors of the Mean, and Skewness Values

Variable	M	SD	SEMean	Skewness
Student body's ethnic/racial composition				
African American	216.1	231.2	24.79	1.42
European American	99.4	164.6	17.75	3.24
Hispanic/Latino	336.4	287.6	31.38	1.27
Student body's linguistic composition				
Native speakers of Spanish	253.1	248.6	27.12	1.41
Monolingual native speakers of English	360.5	244.4	26.82	1.06
Classified as LEP/ELL	130.7	127.2	13.72	1.84
Student body's family socioeconomic status				
Unemployment level	293.5	249.2	27.03	1.21
Public assistance dependence level	315.9	250.0	26.80	1.04
Fully subsidized lunch eligibility level	404.8	252.0	27.50	0.66
Subsidized lunch eligibility level (fully + partly)	461.7	276.1	30.31	0.59
Note. *N* = 83–87 schools. The figures in this appendix are based on the variables measured in counts.				

Copyright 2001 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Sherman Dorn
University of South Florida

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
hmwkhelp@scott.net

Thomas F. Green
Syracuse University

Alison I. Griffith
York University

Arlen Gullickson
Western Michigan University

Ernest R. House
University of Colorado

Aimee Howley
Ohio University

Craig B. Howley
Appalachia Educational Laboratory

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Benjamin Levin
University of Manitoba

Thomas Mauhs-Pugh
Green Mountain College

Dewayne Matthews
Western Interstate Commission for Higher Education

William McInerney
Purdue University

Mary McKeown-Moak
MGT of America (Austin, TX)

Les McLean
University of Toronto

Susan Bobbitt Nolen
University of Washington

Anne L. Pemberton
apembert@pen.k12.va.us

Hugh G. Petrie
SUNY Buffalo

Richard C. Richardson
New York University

Anthony G. Rud Jr.
Purdue University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Jay D. Scribner
University of Texas at Austin

Michael Scriven
scriven@aol.com

Robert E. Stake
University of Illinois—UC

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)

María Beatriz Luce (Brazil)

Departamento de Investigación
Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Maria Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

other vols. | abstracts | editors | board | submit | comment | subscribe | search

This article has been retrieved 2407 times since January 15, 2001

other vols. | abstracts | editors | board | submit | comment | subscribe | search

Education Policy Analysis Archives

Volume 9 Number 2 **January 15, 2001** **ISSN 1068-2341**

A peer-reviewed scholarly journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2001, the **EDUCATION POLICY ANALYSIS ARCHIVES**.
Permission is hereby granted to copy any article
if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Testing Times: A School Case Study

Ivor Goodson
University of Rochester (USA)
&
CARE
University of East Anglia (UK)

Martha Foote
University of Rochester (USA)

Abstract
A highly successful, innovative and creative alternative to traditional education is confronted by the demands of contemporary standardized accountability. The account here is a chronicle of the resistance of a particular school, the Durant School, to the global changes that would destroy its local ecology—a school whose fight against the imposition of state standards and mandated tests has been

> a fight to preserve its integrity, its mission, and its autonomy.

Picture this: a public urban high school conceived in the late 1960s as an alternative to the traditional education and hierarchical structure of most city schools. A school that has not only upheld this unique educational and social vision through its 30-year history, but is deemed successful in terms of its high attendance and college acceptance rates, as well as its low dropout and suspension figures. A school whose 200 students—African-American, White, Latino/a, and Asian-American—choose to enroll there because of this unique vision and high success, and whose teachers choose to work there because they know the school affords them the freedom and respect to realize their innovative educational beliefs. A school that is frequently described by teachers, students, and parents alike as a community, a family even, due to its non-hierarchical structures and close, supportive relationships.

Moreover, these judgments of success are not made only by those involved in this school. The city's mayor recently commented on the school's achievements in a letter to the state education commissioner, noting that the school's "success rate in graduating at-risk students is approximately 20 percent higher than the City School District's average rate." In addition, the school "boasts some of the District's highest attendance rates, highest SAT scores, lowest suspension rates, and lowest dropout rates." The mayor concluded that this school's "non-traditional, yet rigorous process for demanding accountability and assessing knowledge serves its students well." (Note 1) This then is a school that has not only kept its unique vision alive, it has also passed the tests of a school's success that have been set over its thirty years.

Yet, what happens when this school, an oasis of non-traditional practices, is confronted in this current era of educational accountability by an entirely different vision of what a successful school should be? A vision embodied in newly mandated state standards and standardized tests? A vision that, in fact, parallels the over-standardized, over-tested types of schools which the school's original founders turned their backs on 30 years ago in their search for a successful alternative? One would common-sensically expect that any form of governance, state or local, would not change "a winning team," but in the new forms of governance, educational success does not exempt schools from systematic new forms of interference.

In the new regimes of governance in education, control of education is passing from the trusted coalitions of teachers, students and community that have been painstakingly developed in schools such as this. In a more general sense, control is passing from internal educational agents and student and parental communities towards external forces representing a different range of interests. (Note 2) Lobbying efforts by corporations and industrial interests impinge hugely on the judgments of politicians and state education commissioners. These forces drive educational governance in wholly new directions. New patterns of external and symbolic control typically focus on testing, transparency, and accountability. Whilst understandable in principle, in reality such methods often collide with the delicately constructed ecology of school life. As such globalization wreaks environmental havoc in the world generally, so, too, can its specific effects in schools grievously damage the local ecology of an educational environment.

This account, then, is a chronicle of the resistance of a particular school, the Durant School, to just those global changes that would destroy its local ecology—a school whose fight against the imposition of state standards and mandated tests has been a fight to preserve its integrity, its mission, and its autonomy. In other words, it has been a fight both to survive and to defend a different, many would say more humane, vision of schooling.

Before we examine this school more closely, it is important to step back a moment and briefly contemplate a key argument for the standards movement: that the definition and prescription of higher standards will improve our failing schools. Though many dispute the notion that state-mandated curricula imposed in a top-down fashion and policed through the use of high-stakes, standardized exams will improve schools, we need to ask different questions. What will the standards movement do to our *successful* schools? Why must they comply with decrees and edicts pertaining to the content of their curricula when their graduates have a proven record of success in both college and the workplace? Why must their students submit to a battery of paper and pencil exams that supposedly demonstrate academic competency when this competency is already demonstrated by their post-graduation performances, let alone their classroom achievement? [And, we might add, why should the focus be only on strictly academic intelligence when more and more business gurus—the very people often influential in the standards movement—are stressing the crucial importance of social and emotional intelligence?]

The reply from standards advocates has been that if a school is already successful, then the standards and their accompanying tests should amount to nothing more than a few hours out of a student's life to sit for the requisite state exams that she/he will undoubtedly pass if the school is, indeed, of high quality. Such a response starkly exposes the narrow and limited perspective of what many standards advocates believe education is all about: a circumscribed set of skills and myriad facts that can be regurgitated onto a paper and pencil exam in a pressurized testing environment. It is this perspective that the non-traditional Durant School has been fighting in recent months. Not surprisingly, since the school was set up deliberately to alleviate problems generated by a previous era of educational thinking of precisely this kind.

Located in a small, industrial city in the northeast section of the US, the Durant School first faced the possibility of new state standardized exams in 1996. It was in April that year that the state's commissioner of education announced the adoption of a series of five standardized exams— in five different content areas—to measure the attainment of the state's new higher standards by high school students. The passage of all five exams would be mandatory for graduation, and no public high school student would be exempt. Though the exams would be gradually phased in so as to give teachers and students time to prepare, the Durant School was acutely aware of the immediate, and deleterious, impact of these mandates on its program. Specifically, in order to prepare its students for these exams, the school would have to begin both providing courses that specifically addressed the content of these new state standards and preparing students to take standardized exams. Both these practices are antithetical to the school's philosophy that students should have opportunities to learn in-depth in areas of their own interest, and that this learning is best demonstrated through presentations, portfolios, and long-term projects, or in other words, through performance-based assessments. In an attempt to preserve its integrity, an exemption from the state mandates was imperative.

In the summer of 1997, the Durant School applied for a variance from the state exams, maintaining that it upheld and even surpassed the broad state standards. [It is important to note that there are two sets of standards at play in this struggle—the broad state learning standards that address the development of cognitive skills, and the narrow content standards for the different subject areas.] The school asked that instead of exams, it be allowed to continue to evaluate the students' attainment of the broad learning standards through its own performance-based assessments, especially as these very same assessments had recently been publicly commended by the state as a model for high schools to emulate. To its great shock, the state denied the request, maintaining that any alternative assessments to the state exams had to be externally developed; individual schools' assessments could no longer be trusted to ensure high standards. This rejection illustrates just how dramatically the educational and ideological climate has been transformed in the past decade. Performance-based assessments and local control have been knocked from the vanguard, usurped by standardized tests with their scientific claims of "objective" reliability and validity, delivered by bureaucrats from "on-high." However, the Durant School did not surrender its principles so easily: the fight had only just begun.

Throughout the 1997-1998 school year, the principal of the Durant School maintained contact and eventually joined forces with a group of non-traditional high schools in the state, most of which are located together in another city, nearly 400 miles away. These schools were also fighting the state exam mandates, maintaining that their performance-based assessments not only upheld their missions and programs, but were also valid measures of the broad state standards. This union of schools, which now included the Durant School, decided to apply for a group waiver from the exams. However, rather than rushing forward with the request, they thought it best to take their time and build as strong a case for their alternative assessments as they could.

While this group effort was underway, the Durant School, cautious that the state might turn down the group waiver as well, began to examine other possible strategies to circumvent the testing mandates. Charter schools was one idea, and in the fall of 1998, during their biweekly school based planning team meetings, staff, students, and parents discussed together this possibility as a way to preserve the Durant School's autonomy. Though the idea was appealing to some, there was also strong philosophical opposition to such a move, especially regarding the siphoning of public school funds for these schools and their use by the religious right. Later, when it was discovered that charter school students would still be required to pass the state exams to graduate, the idea became moot. During this same period, there was also talk about granting GEDs in lieu of state diplomas. Yet, again, there were grave concerns, especially that such a move would bar future education or job opportunities to Durant School graduates and be publicly perceived as a retreat from quality learning.

While the development of internal strategies for maintaining the school's autonomy and integrity was crucial, the school realized that these strategies alone were not enough, that a public relations campaign was also essential in a successful fight against the state standards mandates. Therefore, as the internal strategies were discussed and debated in the weekly staff and biweekly school based planning team meetings, the Durant School began to pursue several avenues of gaining public support for the school, and consequently, its request for a variance from the state exams. Heeding the advice of a sympathetic member of the city's board of education, the principal and staff enlisted parents, a.k.a. "voters," as lobbyists to advocate for the school. A special meeting was convened in November 1998 for staff to talk with a group of responsive parents about the threat these exams posed to their children's education. These parents in turn offered to organize and attend meetings with members of the board of education and the schools' superintendent to enlist their support. Also, the school's community board—a board consisting of staff, parents, students, and community supporters of the Durant School—decided to organize and sponsor a local conference, open to the public, on the effects of the state exams on student learning.

Meanwhile, the school also turned to the media, especially the local daily newspaper, to publicize its plight. The principal's guest editorial on the negative effects of the state exams on the Durant School was published in mid-November, followed by an in-depth article on the school a few days later. When the same newspaper then published its own editorial claiming that the school could both maintain its program *and* prepare its students for the state exams, an English teacher in the school swiftly responded. In his published letter, he chastised the editorial board for its lack of evidence that the school could do both, indicating that it had not adequately researched the issue. Aside from the daily newspaper, the school also turned to a local radio station for public outreach. Soon the principal, a parent, and a psychology professor from a local university [and a Durant School

Community Board member] appeared together on a talk show to discuss the testing mandates and their effects on learning.

It was also in November 1998 that a math teacher suggested during a school based planning team meeting that the school contact state legislators in an effort to gain their support. His reasoning was that even though the commissioner of education and his board had set the state exam policy, the legislators were the ones in charge of implementation. Following this suggestion, staff, parents, students, alumni, and Community Board members began to write letters to local state legislators, asking for support of the variance. The school also began to solicit the support of business leaders who could, hopefully, influence the state politicians and education leaders.

The public relations campaign continued to gain steam through the winter of 1999. The principal devoted several hours each day drumming up support for the variance request, arranging meetings with political, business, and state education leaders, and seeking public opportunities to spread the word of the harmful effects of the standards mandates on the school. Two parents in particular consistently worked on these efforts with him; the supportive school board member offered strategic advice; and various staff, students, parents, alumni, and Community Board members also volunteered. Staff and school based planning meetings, as well, were filled with regular discussions on the efforts to secure the variance from the state tests. The fight had gained a preeminent position in the school's day-to-day operations, and though staff expressed much stress as a result, they were unwilling to capitulate to the standards mandates.

In February the community board-sponsored conference on the state standards and testing was held. Approximately 100 persons heard Monty Neill, the executive director of the National Center for Fair & Open Testing, give an impassioned keynote address, and lively debate among local and state educators ensued throughout the evening. This event, covered by local television, radio, and newspaper media, was coincidentally followed the next day by a regional hearing on the standards, sponsored by the state education department. Several members of the Durant School community testified, and according to the principal, the students' personal stories of their educational experiences had a profound effect on one member of the commissioner's board, who publicly stated afterwards that she would support a waiver for the school. Buoyed by these small steps, the school pressed on, and more meetings were held with political and educational leaders throughout the spring. Even when support was not secured, the principal was pleased that at least the standards and testing mandates had been raised publicly as an issue that merited deep critical consideration, and that the Durant School had put the word out.

By June 1999 significant local support for a variance had been attained. The superintendent of the city schools, assured that the alternative assessments in the group waiver were, in fact, aligned with the broad state learning standards, had quietly signed on. The board of education, in turn, passed a resolution of support for the waiver, and even the editorial board of the daily newspaper changed its position and came out in favor of a variance for alternative schools. A number of local legislators had responded to the school's requests for support with letters to the education commissioner, asking him to grant the school a variance as well. There was a greater sense of optimism that a variance really was within reach, and that the school's integrity could be preserved.

It was also in June that the Durant School began to lobby the legislative chairs of the joint state education committee, an association that proved especially advantageous in the coming months. The principal had always maintained that if the state education department and the education commissioner did not approve a variance, then special legislation was another possibility. Thus, when the joint legislative education committee announced a June hearing in the state capital to examine the impact of the standards mandates and testing on schools, the principal welcomed the opportunity to make the case for the waiver and gain support for the Durant School's plight. After some preliminary strategy meetings in the weeks before the hearing, about a dozen Durant School representatives—students, staff, parents, Community Board members, and alumni—traveled over 200 miles by rented van to testify. Several other representatives from the alliance of schools seeking the group variance testified as well; and by the day's end, the committee chairs expressed sympathy for the variance request, especially as the students' testimonies to these schools' positive effects on their lives had been, in the chairs' opinion, so persuasive.

Summer 1999, though slower-paced, did see two significant developments in the fight: the mayor wrote a letter to the education commissioner in support of the variance, and a majority of the local legislators signed a pro-variance petition, also addressed to the commissioner. However, as the new school year commenced in September, the cautious optimism in the school began to wane. A ruling on the group variance, now formally submitted, remained pending, and teachers and students expressed deep feelings of anxiety and frustration as they awaited a decision. The education commissioner, they observed, seemed more intransigent than ever as he adamantly, and frequently, proclaimed in the media that there would be no retreat from the state standards—an ominous sign, they believed, for the variance. This apprehension only increased as the missives from the state education department consistently emphasized that the only viable alternative assessments to the state exams would be other externally developed tests. Performance-based assessments, it seemed, were not even considered an option. Despite this pessimism, the Community Board did sponsor another conference at the school on the effects of the standards mandates in an attempt to educate, and galvanize, the public. However, turnout was poor, and several in the Durant School community interpreted this low attendance as an indication that the standards had already been accepted as a fait accompli. They also despaired any prospect of a statewide opposition movement. Still, a letter writing campaign, organized by a parent, was launched to intensify the pressure on political and educational leaders, and the school continued to wait anxiously for an official ruling on the variance.

It was during this bleak period that a group of Durant School students, disgusted by the fact-filled, rote learning of their newly mandated history class, decided to act. As second-year students they had previously experienced the pleasure of the school's learner-centered classes, and they were outraged by the difference in this class, especially as it was instigated by the state standards. When the school sent representatives to speak at a regional joint legislative education committee hearing, this time only 100 miles away, about 20 students voluntarily attended, either to testify or show support. Again, the committee was deeply impressed by the students' spirit and pride in their school, and a legislative aide privately predicted that the waiver would be granted. This development, combined with reports that other students from the alliance of schools had also made a strong impression at their regional hearing, helped re-energize the fight. In addition, the staff began to work monthly with a volunteer business consultant on ways to focus their energy in fighting the mandates and gaining support for the variance.

In December 1999 the state's official response to the variance request began to take shape as the Assessment Panel of the State Education Department granted the alliance of schools a hearing in which to present their assessments. The alliance, in turn, solicited six nationally-known educational leaders, and friends of the alliance schools, to make the presentation. Not only did the alliance believe that these leaders, who also served on the alliance's performance assessment review board, would present a strong and convincing case, they also believed, according to the Durant School principal, that their prestige would lend political weight to the variance request. The night before the hearing, the six leaders gathered with several representatives from the alliance schools to discuss strategy and outline the presentation. At the two-hour hearing the following day, the six argued the case for the variance, answered questions from the committee, and defended the quality of the alliance's system of assessment. When the hearing concluded, a press conference, arranged by the alliance, was held in which the presenters attested to the urgent need for the variance.

That same day, the state's Assessment Panel issued its recommendation to the education commissioner: only a partial variance be granted, limited to the schools covered by a previous variance from state exams [this limitation excluded the Durant School], and good for only one year. When this recommendation was made known, the Durant School immediately intensified its campaign. The principal and several parents implored the school community to call and write letters to the legislative education committee members, urging them to request a full variance for the school from the commissioner. The community responded with a flurry of activity. The alliance, in turn, scheduled meetings with the education committee chairs to ask them to lobby the commissioner for the full variance as well. Finally, the day of reckoning arrived at the end of January 2000. The commissioner, following most of the panel's recommendations, issued a partial variance through the 2000-2001 school year, limited to the alliance schools in the previous variance. However, he did approve an extension of the variance to any remaining alliance schools that could demonstrate they had met the criteria of the alliance. This extension provision kept the Durant School's hopes alive, as they were certain of having already met all the criteria. By March, after the school had submitted proper documentation, the commissioner ruled that the Durant School was also covered under the temporary waiver. Significantly, the daily newspaper reported the story on the same day as it published an in- depth feature article on the Durant School in its series on the city schools, an article that had been actively solicited by the principal.

As of March 2000, the partial variance is only a partial victory. Keeping in mind that the five exams are being gradually phased in, this year's seniors are exempt from their only required exam, specifically English Language Arts. This year's juniors, however, must take, and pass, the English Language Arts exam to graduate, though they are exempt from the requisite state math exam, the second exam to be phased in. The current sophomores and freshmen have no exemptions -- they must pass four and five exams, respectively, in English language arts, math, world history, American history, and science, as all five mandated exams will be required of the Class of 2003.

Despite the commissioner's ruling, the fight is not over. The Durant School, both alone and with the alliance, continues to devise strategy, lobby for supporters, and struggle to attain a full and complete variance. The activist spirit in which this school was created is alive and well, and it offers hope, 30 years later. In particular, it offers a model of how a socio-political process of advocacy and campaigning can turn the juggernaut of external forces in ways that benefit the educational endeavor. For, contrary to the position of the standards movement proponents, educational success, as epitomized by this school, is indeed attainable through the efforts of internal agents—coalitions of teachers, students, and parents. These are the only agents who can truly know a particular school, thus possess the insight to determine what makes it "succeed" in the most profound sense of the word, and not as a simplistic reduction to a standardized test score.

Notes

1. Mayor's letter to State Education Commissioner, June 28, 1999.
2. Goodson, I. (Forthcoming) Social Histories of Educational Change Theory in *The International Journal of Educational Change.*

This report originates from a research study entitled "Change Over Time," funded by the Spencer Foundation. Our thanks for their continued support.

About the Authors

Ivor Goodson
University of Rochester (USA)
CARE, University of East Anglia (UK)

Email: ivor@troi.cc.rochester.edu

Ivor Goodson is Co-Director of the "Change Over Time" research project. He is Professor of Education at both the University of East Anglia and the Warner Graduate School at the University of Rochester, USA. He has been Director of two research units, most recently at the University of Western Ontario where he has directed a wide range of research projects on computer education, teachers lives and careers, case histories of school and curriculum, environmental education and racial ethnocultural minority teaching. Among his books are *School Subjects and Curriculum Change, The European Dimension and the School, The Making of Curriculum: Collected Essays, Studying School Subjects, Studying Teachers' Lives, Biography, Identity and Schooling, Studying Teachers' Lives* and Subject Knowledge: Readings for the Study of School Subjects , *The Changing Curriculum: Studies in Social Construction*. His books have been translated in Spain, Sweden, Portugal, Germany and Brazil, and are forthcoming in Japan, China and Finland. He is founding editor of the *Journal of Education Policy* and the editor of book series for Falmer Press, the Open University Press and Teachers College Press. He is the author of "The Devil's Bargin: Educational Research and the Teacher," *Education Policy Analysis Archives, 1*,(3).

Martha Foote
Research Officer
Project "Change Over Time"
Warner Graduate School
University of Rochester

Email: martha_foote@hotmail.com

Martha Foote was a classroom teacher for ten years before becoming a PhD student in Teaching and Curriculum at the Warner Graduate School of Education and Human Development at the University of Rochester.

Copyright 2001 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Sherman Dorn
University of South Florida

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
hmwkhelp@scott.net

Thomas F. Green
Syracuse University

Alison I. Griffith
York University

Arlen Gullickson
Western Michigan University

Ernest R. House
University of Colorado

Aimee Howley
Ohio University

Craig B. Howley
Appalachia Educational Laboratory

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Benjamin Levin
University of Manitoba

Thomas Mauhs-Pugh
Green Mountain College

Dewayne Matthews
Western Interstate Commission for Higher Education

William McInerney
Purdue University

Mary McKeown-Moak
MGT of America (Austin, TX)

Les McLean
University of Toronto

Susan Bobbitt Nolen
University of Washington

Anne L. Pemberton
apembert@pen.k12.va.us

Hugh G. Petrie
SUNY Buffalo

Richard C. Richardson
New York University

Anthony G. Rud Jr.
Purdue University

Dennis Sayers
Ann Leavenworth Center for Accelerated Learning

Jay D. Scribner
University of Texas at Austin

Michael Scriven
scriven@aol.com

Robert E. Stake
University of Illinois—UC

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

U.S. Department of Education Brigham Young University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

other vols. | abstracts | editors | board | submit | comment | subscribe | search

This article has been retrieved 488 times since January 22, 2001

other vols. | abstracts | editors | board | submit | comment | subscribe | search

Education Policy Analysis Archives

Volume 9 Number 3 **January 22, 2001** **ISSN 1068-2341**

A peer-reviewed scholarly journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2001, the **EDUCATION POLICY ANALYSIS ARCHIVES**.
Permission is hereby granted to copy any article
if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education.*

Impact of U.S. Overseas Schools in Latin America on Political and Civic Values Formation

John J. Ketterer
Jacksonville State University

George E. Marsh II
University of Alabama

Abstract

This study focuses on the attitudinal outcomes of schooling in American Overseas Schools in Latin America with respect to democracy and citizenship, the formation of views about the United States, and student attitudes about the American international school.

Introduction

The American democracy is the oldest in the world and the promotion of democracy has been a central focus of U.S. foreign policy since World War I. The evolution of Latin American nations towards democratic models of governance during the 1980's was trumpeted as a diplomatic triumph. The argument has even been made, prematurely perhaps, that the historical process of the selection of an ideal model of governance has ended and that the democratic model has emerged triumphant (Fukuyama, 1992). Although the decade of the 1990's saw some regression in this process, virtually every nation from Mexico to Brazil has attempted to develop democratic institutions. Many of these "experiments" are yet in their infancy and all of them depend upon the values and ideals of leaders who will be elected to key offices in the future. Diamond (1993) documents the importance of educational institutions; he mentions the "international diffusion of values and beliefs" which may occur through practices which occur within "democratizing institutions" (p. 421). He observes that

> Culture springs from history, tradition, and collective myths, and is also forged and reproduced through a variety of institutional settings in which norms are learned, beliefs generated, and values internalized. Prominent among these settings are, of course, the family and the school… [which may] contribute to significant change over time. (p. 412)

It is a little known but important fact that a significant number of political and business leaders in Latin American nations have been educated in American Overseas Schools (AOS), and many enter American universities after successful completion of an American high school education in an overseas school. Bilingual and infused with the values implicit in U.S. pedagogy, these young people become the mayors, judges, industrialists, journalists, cabinet ministers, and presidents of their countries. Clearly, the political culture of the United States has profound direct and indirect influences on the attitudes of the future leaders of Latin America. There have been no studies focusing on the attitudinal outcomes of students in American schools overseas.

The AOS schools are essentially American high schools in Latin America. Typically, these schools offer a traditional, college preparatory American high school curriculum. Unlike AOS schools in other regions of the world, the AOS in Latin America frequently incorporate host country languages and national curricula in the school model. However, American citizens trained and certified in American universities serve as principals and certified American teachers deliver the central elements of the curriculum. With the fiscal and technical support and guidance of the Office of Overseas Schools of the U.S. Department of State, most of these schools have achieved accreditation by the Southern Association of Colleges and Schools (SACS), the entity which accredits institutions in the United States from Texas to North Carolina. (The Office of Overseas Schools is staffed with a Director and six Regional Education Officers, each assigned oversight of a geographic region. The Director of the Office is Dr. Keith D. Miller (millerkd2@state.gov). The web site of the Office of Overseas Schools may be found at http://www.state.gov/www/about_state/schools/ofront.html.) Many of the AOS schools have a long history, such as the American School Foundation (ASF) of Mexico City, which has operated an American-type school with an American curriculum for over 100 years. Half of the ASF students enroll in colleges abroad, predominantly in the United States. Although these schools were originally established to educate the children of American citizens who lived with their families in Latin America (as part of the diplomatic corps or the international business community), that mission has clearly been altered by economic and political factors. Orr (1974) observed that the schools "exemplify the valuable qualities and merits of a democratic educational system" and serve as a "living example of American community democracy" (p. 10). He declared that "The success or failure of the U.S.A., both internally and as a model, will be directly related to the effectiveness of education and schooling" (1981, p. 2). Conlan (1982) spoke of the AOS schools as "isomorphic embassies."

As the world economy changed over the years, host-country children in Latin America were increasingly drawn to American schools where they could learn English. The downsizing of the U.S. diplomatic corps and a concomitant "nationalization" of the work force in the international business community accelerated this demographic change in the 1970's. American schools have retained a "U.S." identity through the networking of regional educational associations, greater use of the Internet than comparable schools in the continental United States, and the recruitment and training of U.S. teachers who already possess advanced degrees from U.S. universities. American history, civics, and literature are central to the curriculum. Host-country students, from Mexico to Brazil, who graduate from these schools receive the American high school diploma (commonly they also receive the host country diploma, or "bachillerato"). Most plan to attend U.S. universities, either as undergraduates or for graduate study, and later return and assume responsible positions in their homelands.

Purpose

The unique role that a U.S. education plays in the career planning of future Latin American leaders has not been examined, although it has been a subject of comment. AOS schools directly influence the development of the values and attitudes of many Latin American leaders. The purpose of this research was to assess the political attitudes of 12th grade students attending 12 AOS schools in 8 countries. Three distinct groups of students were targeted in this study: American citizens, Host Country Citizens, and Students who were citizens of some third country (children of parents who form part of the international diplomatic or business community). The supposition that American Overseas Schools contribute to the formation of positive values of democratic participation and civic service should be investigated. Arguably, the extent to which these schools are in fact promoting these values is a valid measure of the efficacy of the schools themselves.

Research Questions

Three research questions were developed for this study. (1) Is there a significant interaction effect between the independent variables of political region and citizenship on students' attitudes? (2) What is the relationship between the length of time a student is enrolled in an American school and the development of positive attitudes? (3) Is there an attitudinal difference with respect to gender on these measures?

Method

Subjects. The subjects of this study were 695 12th grade students representing 21% of the approximately 3,200 12th grade students attending AOS schools in 4 geographical and political regions: Mexico, Central America, Spanish-speaking South America, and Brazil. The schools were distributed among the following countries: Mexico (3), El Salvador (1), Guatemala (1), Paraguay (1), Ecuador (1), Argentina (1), Peru (1), and Brazil (3). U.S. citizens represented 15.3% of the sample and host country nationals represented 68.2% of the sample. The other 16.5% was accounted for by third-country nationals, pupils who were not American citizens or citizens of the countries where they attended schools.

Instrument. The survey instrument, Attitudes toward Democracy (ATD©), consisted of 40 Likert-type items based on a 5-point rating system ranging from strongly agree to strongly disagree. The items were associated with three categories, concerning (a) attitudes about democracy, citizenship and service, (b) attitudes toward the United States, and (c) attitudes about the role of school. The first scale combined the two aspects of responsible democratic participation, rights and obligations (People for the American Way, 1989). The second scale measured student attitudes about the U.S. government and overall attitudes about the people of the United States. The third scale assessed student attitudes about the role of the school in their social and political formation.

The instrument had high overall reliability (Cronbach Alpha

= .85) and the three scales individually yielded alphas of .85, .70, and .68, respectively. The ATD instrument was mailed to the directors of the 12 schools and administered under the supervision of certified teachers according to a set of standard instructions.

Results and Discussion

An ANOVA revealed a significant interaction [$F(6,683)=2.41$, $p<.05$] between the variables of citizenship and political region on Scale 1, attitudes toward democracy and citizenship. Citizens of Mexico, Central America, and Brazil had significantly more positive attitudes on this scale than their counterparts in Spanish-speaking South America. U.S students in Brazil had significantly less positive attitudes than U.S. students in Mexico. Host country students in Brazil had significantly more positive attitudes than U.S. students in Brazil.

There was no significant interaction between the two classes of independent variables on Scale 2, although there were significant main effects in both areas. Table 1 shows the ANOVA for Scale 2, attitudes toward the United States. Significant differences were found between the attitudes of U.S. citizens and the other two groups. Attitudes of the host and third country pupils were significantly more negative, and the mean response of both groups was to the negative side of the scale.

Table 1
Analysis of Variance for Scores on Scale 2: "Attitudes Towards the United States"

Source of Variation	Sum of Squares	DF	Mean Squares	F	*p*
Main Effects:	1310.52	5	262.10	5.88	<.000
Citizen	674.91	2	337.45	7.57	.001
Region	380.66	3	126.89	2.85	.037
Interaction: Citizen X Region	385.45	6	64.24	1.44	.196
Explained	1977.30	11	179.75	4.03	<.000
Residual	30452.04	683	44.59		
Total	32429.33	694	46.73		

The ANOVA revealed a marginally significant interaction [$F(6,683)=1.94$, $p<.10$] between the independent variables of political region and citizenship for Scale 3, attitudes about the role of the school. Interestingly, host country students in Mexico were shown to have significantly more positive attitudes about the United States than host country students in the other regions.

The length of time enrolled in the AOS school had no relationship to the development of positive attitudes about the United States (correlation = -.006; p=.89). However, student attitudes on Scale 1 (Attitudes about Democracy, Citizenship and Service) demonstrated a positive correlation (correlation = .143; p<.001). Similarly, with respect to Scale 3 (Attitudes about the School), student attitudes were found to satisfy the statistical test for significance (correlation = .087; p=.02). However, it must be noted that these correlations, given the large sample size, are so close to zero as to provide little evidence of a causal relationship, even if they could be so interpreted.

To measure the relationship between the variables of gender and the mean student responses of each of the three scales, *t*-tests were calculated for the independent samples. A significant difference (*t*=-3.90, df=693, p=<.000, 2 Tail Sig.) was found on Scale 2, attitudes about the United States. Female students had significantly more positive attitudes than male students about the United States.

Although the data revealed a large number of interesting relationships and circumstances, a summary of the main findings follows:

1. Twelfth grade students in AOS schools who are citizens of South American countries possess extremely negative attitudes about democracy and citizenship.
2. U.S. citizens who are 12th grade students in AOS schools in Brazil are negative about democracy and citizenship.
3. International and host country students in all of the Latin American AOS schools are extremely negative about the United States. U.S. 12th grade students were predictably more upbeat.
4. Mexican students in the 12th grade in AOS schools expressed significantly more positive attitudes about the United States than their counterparts in other regions.
5. Female 12th grade students in the AOS schools expressed more positive attitudes about the United States that the males in the same schools.
6. The length of time a student is enrolled in the AOS school has no clear impact on the development of positive attitudes about democracy, the United States, or the role of the school in the social formation of the student.

Conclusions

The generally negative attitudes about the United States expressed by students throughout Latin America in the AOS schools should be a matter of concern for the U.S. State Department, which oversees these schools. A programmatic approach system-wide to social studies curricula should be considered. If the American Overseas School serves the quasi-diplomatic function of modeling democratic institutional behavior, then educators should focus on developing a model with the express purpose of promoting positive attitudes. It should be noted, however, that at least a portion of the negative response might be age-related, and there is some evidence that with time and maturity these attitudes may improve.

The relatively more positive attitudes of Mexican students may well point to a strategy for improvement of student attitudes in other regions. The AOS schools in Mexico are among the oldest in the world. They are generally viewed as deeply embedded in host country culture. They have traditionally incorporated the Mexican curriculum into the U.S. curricular model as an enriching factor. The fact that Mexican culture has been "included" rather that "excluded" in the structure of these schools may be a factor in the more positive attitudes of Mexican students.

The lack of impact of the time a student spends in the AOS school on the development of his/her attitudes is disappointing. This is yet another indication that school leaders and regional planners should focus on the formation of students' attitudes as a valid formative goal of the school curriculum.

The significant difference between the attitudes about the United States of young women and young men in these schools can only fuel speculation. It may be that the threat of economic competition with the United States is more acute for young men than for young women. We might also speculate about traditional roles of women in Latin America, the attractiveness of U.S. popular culture, and other factors. For the present, this finding must remain an interesting puzzle, although further investigation as to its cause might indicate a path that would lead to general attitudinal improvement.

The findings of this study lead to new and important questions about the role of the school in the attitude formation of students. How should the school model reflect or incorporate the cultural context? Can the curricula of these schools be restructured to improve attitudinal outcomes? The mission of the AOS schools is generally understood to be that of representing a positive model of an effective democratic institution. Because this is the case, the U.S. State Department's Office of Overseas Schools and regional educational leaders should take actions directed at programmatically and systematically addressing that goal.

References

Conlan, A. L. (1982). The contributions of American sponsored overseas schools in fostering, promoting, and maintaining cross-cultural understanding between and among the peoples of the Americas (Doctoral dissertation, The University of Alabama, 1982). *Dissertation Abstracts International, 32,* 4273.

Diamond, L. J. (1993). *Political culture and democracy in developing countries.* London/Boulder, CO: Lynne Rienner Publishers.

Fukuyama, F. (1992). *The end of history and the last man.* New York: MacMillan.

Ketterer, J.J. (1998). The impact of selected American schools in Latin America on the attitudes of students regarding democratic values, the obligations of citizenship and community service, and the United States of America (Doctoral dissertation, The University of Alabama, 1998). *Dissertation Abstracts International, 59*(05), 1413d.

Orr, P. G. (1974). *A research matrix: The American sponsored overseas school*. Buzzard Bay, MA: Center for International Education, Massachusetts State College System.

Orr, P. G. (1981). *Overseas education: Quo vadis and the quid pro quo*. [Fast Reference Series No. 018]. Tuscaloosa, AL: University of Alabama, International Education Associates, Inc.

People for the American Way. (1989). *Democracy's next generation: A study of youth and teachers*. Washington, DC: Peter D. Hart Research Associates, Inc. (ERIC Document Reproduction Service No. ED 324 253).

About the Authors

John J. Ketterer
College of Education
Educational Resources
Jacksonville State University
Jacksonville, AL 36265
Voice: (256) 782-5837
Fax: (256) 782-5959
Email: jkettere@jsucc.jsu.edu

John J. Ketterer, Ed.D. is currently an Assistant Professor at Jacksonville State University, Jacksonville, Alabama. He is a former school superintendent, and worked extensively in overseas assignments. His current research interests include the study of the impact of postmodern philosophy on school policy and programming, bilingual and multicultural education, the role of education in democracy, and distance education.

George E. Marsh II
College of Education
Instructional Technology
The University of Alabama
PO Box 870302
Tuscaloosa, AL 35487-0302
Voice: 205-333-9105
Fax: 205-333-8288
Email: gemarsh@bamaed.ua.edu

George E. Marsh II, Ed.D. is currently a Professor in Instructional

Technology at The University of Alabama, Tuscaloosa. His research interests include epistemological studies and applications to organizational theory and instructional design, distance education, and online delivery systems.

Copyright 2001 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Greg Camilli
Rutgers University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

David D. Williams
Brigham Young University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

other vols. | abstracts | editors | board | submit | comment | subscribe | search

This article has been retrieved 896 times since February 8, 2001

Education Policy Analysis Archives

Volume 9 Number 4 February 8, 2001 ISSN 1068-2341

A peer-reviewed scholarly journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2001, the **EDUCATION POLICY ANALYSIS ARCHIVES**.
Permission is hereby granted to copy any article if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Factors Influencing GED and Diploma Attainment of High School Dropouts

Jeffrey C. Wayman
Colorado State University

Abstract

This study examined correlates of degree attainment in high school dropouts. Participants were high school dropouts of Mexican American or non-Latino white descent who had no degree, a high school degree, or a GED certificate. This study was unique in that it accounted for sample bias of missing data through the use of multiple imputation, it considered students who had dropped out as early as 7th grade, and it was able to include variables found significant in previous research on returning dropouts. Logistic regression analyses identified a parsimonious set of factors which distinguished dropouts who held degrees (diploma or GED) from those who did not. Similar analyses were performed to distinguish participants who had attained diplomas from those who had attained GEDs. It was estimated that 59.2% of dropouts return to obtain high school credentials. School capability, age at dropout, and socio-economic status significantly predicted degree attainment. Presence of children, higher school capability and socio-economic status were associated with GED attainment, while later grade at dropout was associated with diploma attainment. These relationships did not vary by ethnicity, although degree attainment was less likely for Mexican American dropouts. The study concludes that dropping out is not the end of a student's education, and more research should be directed toward returning dropouts. Further, the focus of such research should be expanded to include a more positive and broader range of correlates.

Introduction

Dropping out of high school is a well-documented social problem, and often presents daunting circumstances for adolescents. Dropping out is often associated with delinquency, substance use, and low school achievement (Chavez, Oetting, & Swaim, 1994; Ekstrom, Goertz, Pollack, & Rock, 1986; Elliott, Huizinga, & Ageton, 1985). Further, people without high school degrees often experience lower wages and higher unemployment, and more dependency on welfare and other social services (Catterall, 1987; Rumberger, 1987).

Research also shows that dropping out of high school does not have to be, and is not necessarily, a permanent condition. Estimates of the percentage of dropouts who eventually attain either high school diplomas or General Educational Development certifications (GEDs) have been as high as 44% (Kolstad & Kaufman, 1989). Thus, study of the correlates of degree attainment in dropouts could be an effective tool in reducing the dropout rate, but unfortunately, few studies have been conducted in this area. Balancing the well- developed research on dropout correlates with a research base of return correlates not only provides information on why dropouts gain degrees, but also provides a different perspective from which to augment dropout prevention efforts.

Dropouts Who Return to School Settings

Studies of returning dropouts have examined either dropouts who return to school (Borus & Carpenter, 1983; Chuang, 1997) or dropouts who obtain high school degrees or GEDs (Kaufman, 1988; Kolstad & Owings, 1986; Kolstad & Kaufman, 1989). Studies of this type have compared factors present in returning dropouts to a "typical dropout profile". From the vast amount of dropout literature, these studies have been able to identify factors associated with dropping out and have analyzed variables identified in this profile, hypothesizing that those dropouts who do not fit the profile are more likely to return to high school.

This body of research is not yet sufficiently developed to identify a complete picture of why dropouts return to school settings, although some factors appear to be fairly robust. For instance, achievement test scores were found in all studies reviewed here (except Borus & Carpenter (1983)) to be positively related to return for more education. Early dropouts are less likely to return, as shown by all the studies except Kaufman (1988), which did not include this variable.

Nonetheless, the sparsity of studies on returning dropouts have left many questions as to other variables affecting return. Ethnic effects are an inconsistent mix in these studies, and other factors, such as socio-economic status, are significant in some studies and not in others. Further, questions remain as to the effects of sampling on significant relationships identified – none of these studies were able to consider dropouts who left school before 10th grade, and none were able to estimate effects due to inability to longitudinally follow each participant in the sample.

Previous research has laid the foundation for knowledge regarding degree attainment in high school dropouts. However, such research should be extended and clarified. The next logical step is a study which can pull together significant factors found in previous studies and present estimates which infer to the entire population of dropouts. The present study will address these issues.

The Present Study

The present study examines Mexican American and non-Latino white dropouts who have gained high school diplomas, GEDs, or neither, identifying factors which are associated with attainment of high school credentials. In doing so, this study will address several important problems left unsolved by previous studies on returning dropouts.

First, the present study accounts for bias introduced by dropouts who did not respond to the second wave of data collection. Longitudinal dropout studies naturally suffer from an inability to resurvey each and every dropout. However, each of the reviewed studies conducted analyses on only those dropouts who were successfully followed up. Such treatment of missing dropouts assumes that the dropouts who remained in the study are similar to the ones who did not, an assumption which leaves the study vulnerable to sample bias. The present study, through the use of multiple imputation, accounts for bias caused by missing data.

Second, previous studies were limited to participants who dropped out in tenth grade or later. Although against the law in many states, the truth is that many students leave school before age sixteen. The present study is able to consider students who dropped out earlier than tenth grade – some as early as seventh grade. Inclusion of these students, along with the estimation of missing data described above enables the present study to estimate return correlates for the full dropout population.

Third is the breadth of variables studied in this work. Previous studies independently drew upon factors known to be associated with dropping out and did not purposely examine variables shown to be significant in previous studies of returning dropouts. Therefore, it is not clear whether identified significances are due to omission of other important factors. To truly assess the significance of factors associated with returning dropouts, these factors should be considered in tandem. The present study addresses this need, as all variables considered were chosen based on their significance in previous return research.

Fourth, only Kolstad and Kaufman (1989) considered diploma attainment and GED attainment separately. The present study will also discern differences between students with no degree, students with diplomas, and students with GEDs.

Method

The data for this study were gathered as part of a longitudinal project designed to study substance use and other correlates of high school dropout among Mexican American and non-Latino white dropouts. The sample for this study consisted of Mexican American and non-Latino white adolescent dropouts from three communities in the southwestern United States: a city with 400,000 people, a mid-sized town with 90,000 people, and a small town with 30,000 people. Dropouts were defined as students in grades 7 – 12 who had not attended school for at least 30 days, had not transferred to another school, were not being home-schooled, and had not contacted the school system about re-admission. This definition is more stringent than that recommended by Morrow (1986), whose standard definition of a dropout calls for a period of unexcused absence from school of two weeks or more. The adoption of a period of absence of one month or longer provides a sufficient period of time to ensure that youth are, in fact, high school dropouts.

Potential participants were adolescents from dropout lists provided by school personnel in the aforementioned communities. Once they were identified and contacted, refusal rates were low (4 – 6%), so the resulting sample is a random sample from the population of dropouts from these three communities. Results from this study will be inferred to the population of Mexican American and non-Latino white dropouts in the United States. Although the sampling frame is limited geographically, previous results published from this data set have been comparable to other studies of high school dropouts (e.g., Chavez, Oetting & Swaim, 1994; Chavez, Deffenbacher, & Wayman, 1996). Therefore, inferring to this population from the present sample is appropriate.

Measures

All survey items used in this study were embedded in a larger survey which took approximately one and a half hours to complete. Nearly all surveys were completed in English, with less than 1% completed in Spanish.

Dependent variable. Graduation from high school, possession of a GED, or no degree attainment were based on self-report measures.

Demographic information. Ethnicity was determined from school records and was double-checked by field workers with the participant. Gender and socio-economic status (SES) were based on self-reports from a demographic section of the initial survey. SES was a composite measure of the following items: education of mother, education of father (possible responses of 6^{th} – 12^{th} grade, 1 – 4 years of college, or 5 or more years of college were coded as 6 – 17), "do your parents have good jobs" (possible responses "they do not work", "poor", "not too good", "good", or "very good"), "what is your parents' income" (possible responses were "very low", "low", "average", "high", or "very high") and "does your family have enough money to buy the things you want" (possible responses "almost never", "some of the time", "yes, most of the time", or "yes, all of the time"). Since these items were not uniform in range of possible answers, responses were standardized before being summed to create the composite. The Cronbach alpha reliability of this scale was .65.

Independent variables. Achievement test scores, age at dropout, grade at dropout, and grade point average were obtained from high school records. Achievement test scores were used as a proxy for ability (or "school capability"), which was measured by averaging mathematics, reading and vocabulary scores (Kaufman, 1988) for each participant. Data were collected on achievement tests administered at many times during the participant's school career, but due to inconsistent record keeping, students transferring from districts using different procedures, etc., neither the time frame nor the quantity of test scores was uniform across participants. Thus, the highest available mathematics, reading and vocabulary scores were used. This not only provided consistency, but reduced noise in the test scores as measures of school capability – few students would attain a test score which was a higher representation of their true capability.

Whether the participant had or was expecting children was based on self- reports from the initial survey, as was teacher caring. To assess a participant's feeling of teacher caring, an item asking "how much did teachers care about you during this last year" was included on the survey, with possible responses of "not at all", "not much", "some", and "a lot". Marriage was not used in this analysis because only three of the participants reported being married at the time of dropout.

Procedure

For the first wave of data collection, dropouts were chosen randomly from monthly lists of dropouts, provided by the school district. Field workers, employed by the district and fluent in English and Spanish, first contacted potential participants. After the project was described, potential participants were asked if they wished to be involved. If they expressed interest and were over 18, they completed consent forms. If they were under 18, parents were contacted, the project was fully explained, and written parental consent was obtained. Those who refused were replaced in the sampling frame by another randomly sampled dropout.

Following informed consent, arrangements were then made for an individual administration of the survey. The survey was completed at school or at another public building such as a library, and participants were given as much time as needed to complete the survey. The survey administrator gave participants the survey, answered general questions and helped participants with reading problems, but did not see participant responses. When the survey was complete, the participant put it in a large envelope and sealed it personally. Based on the participant's choice, the survey was mailed to the research office either by the survey administrator or was taken immediately to a mailbox by the participant and survey administrator. These steps assured confidentiality; at no time was an unsealed, completed survey out of the participant's sight. Participants received $25 for completion of the survey.

Accuracy and reliability of data were assured as surveys were subjected to 40 checks for inconsistency or exaggeration (e.g., endorsing a fake drug, claiming daily use of three or four drugs). Only 2% of initial surveys failed either review and were not replaced.

Four years after the first assessment, follow-up of dropouts 18 or older began, with an average time to completion of the follow-up survey of 4.29 years. Follow-up contact was first attempted through the address given at the first assessment. If this failed, staff contacted three people (e.g., parents, relatives, good friends) whom the participant indicated at the time of informed consent would always know where the participant lived. If these efforts failed, public records such as phone books, motor vehicle records, etc., were checked to locate an address. A total of 519 (49%) of the 1071 original participants were successfully followed up. Once the individual was contacted and gave his/her consent, survey administration was parallel to the first administration.

Data Analysis

Multiple imputation. Missing data presented a potential problem in this project, since not all participants had responded to the second wave of data collection. Typically, data such as these are analyzed by using only the cases with fully completed responses in both waves on all relevant variables, discarding incomplete responses. Treating the data in this fashion not only results in a reduction of sample size, but more importantly, implicitly assumes the group of participants who answered all questions to be similar to the group who did not. Should this assumption not hold true, sample bias results. Specific to the present work, analyzing only participants who were followed up presumes these dropouts to have similar characteristics to the dropouts who were not successfully located or who refused to participate. Further, inclusion of only those participants who answered all items would result in a substantially reduced sample size. To address issues of bias and power, multiple imputation was used to account for the missing data in this study (Rubin, 1987; Schafer, 1997). Multiple imputation has been shown to be an appropriate and robust method for estimating missing data in social science settings (Graham, Hofer, Donaldson, MacKinnon, & Schafer, 1997).

In multiple imputation, missing values for any variable are predicted using existing values from other variables. The predicted values, called "imputes", are substituted for the missing values, resulting in a full data set called an "imputed data set". This process is performed multiple times; results from the imputed data sets are combined for the analysis.

Multiple imputation accounts for missing data by restoring not only the natural variability in the missing data, but also by incorporating the uncertainty caused by estimating missing data. Maintaining the original variability of the missing data is done by creating values which are modeled as a function of variables correlated with the missing data and with the causes of "missingness." Random errors from a normal distribution are added to these predicted values to produce the imputed values. Imputed values produced from an imputation model are not intended to be "guesses" as to what a particular missing value might be; rather, this modeling is intended to create an imputed data set which maintains the overall variability in the population while preserving relationships with other variables.

To incorporate the uncertainty associated with estimating missing data, *K* multiple models are drawn from the distribution of plausible models for the population. These models are used to produce *K* imputed data sets. Parameter estimates are then obtained by combining these *K* imputed data sets.

The parameter of interest in the current study is the log odds, denoted by θ in the formulas below. Parameter estimates are computed by averaging the point estimates, $\hat{\theta}_k$, obtained from the imputed data sets thusly:

$$\bar{\theta} = \sum \hat{\theta}_k / K \; .$$

The total variance of $\bar{\theta}$ is given by the formula

$$\boldsymbol{T = W' + (1 + K^{-1})B,}$$

where $W' = \sum W_k / K$, the average of the *K* imputed variances, and

$B = \sum (\hat{\theta} - \bar{\theta})^2 / (K - 1)$, the between-imputation variance of the estimates of θ.

Thus, the total variance of $\bar{\theta}$ is made up of a within-imputation component, W', which estimates the natural variability in the data, and a between-imputation component, B, which estimates uncertainty caused by estimating missing data (Rubin, 1987). Confidence intervals (95%) for θ are given by the usual formula,

$$\bar{\theta} \pm t_{.025, df}(T),$$

with confidence intervals for odds ratios obtained by exponentiating the bounds of the confidence intervals for theta. Degrees of freedom for t-statistics are given by the formula

$$df = (K-1)[1 + KW'(K+1)^{-1}B^{-1}]^{2}$$

Multiple imputation and combination of parameter estimates was performed using the NORM for Windows software package (Schafer, 1999).

Multiple imputation is an appropriate method for treating missing data if correlates of the dependent variable are considered and if the causes of the missing data are measured and available for analysis. To this end, it is important the imputation model is carefully chosen, ensuring that biases introduced by "missingness" are eliminated. The variables which were included in the logistic regression models were necessarily included in the imputation modeling. Also utilized were items correlated with "missingness": location (city or mid-sized community), substance involvement, whether the participant had ever been suspended from school, whether the participant moved into the district from another district, current living arrangements, and whether the participant's family rented or owned their house.

Logistic regression modeling. The research questions in the present study were answered through logistic regression analysis, defining two separate dichotomies as dependent variables – degree/no degree, and diploma/GED. Thus, one set of logistic regression models was estimated to ascertain factors which significantly predict attainment of a high school education (either a diploma or GED) or attainment of nothing. Then, the sample was restricted to participants who have attained a high school education, and models were estimated which distinguish between possession of a diploma versus possession of a GED.

Model selection was performed using a hierarchical backward selection process. In each model, all main effects were examined, along with two-way interactions involving ethnicity, gender and SES (Other interactions were too numerous to examine in one analysis, and no theoretical base was available to justify inclusion or exclusion of particular interactions. The demographic variables ethnicity, gender and SES are the most commonly included variables in return research and are therefore the most pertinent to include in interactions). From this "full" model, interactions were examined separately for significance at the .05 level, using the Wald statistic. The interaction with the smallest Wald statistic was eliminated from the model, then the model was re- estimated with the remaining main effects and interactions. This process was repeated until only main effects and significant interactions remained, if any interactions were significant. If interactions were significant, the main effects supporting these interactions were necessarily retained in the model. The process then was performed similarly for main effects not involved in significant interactions. This process was repeated until the remaining model consisted only of significant factors. These factors were then retained as the most parsimonious set of factors which described the outcome.

For each model, slope estimates (ß's) and standard deviations of slope estimates were obtained by performing a separate logistic regression analysis on each imputed data set. These slope estimates and standard errors were then combined as described in "Multiple imputation" above, producing one set of slope estimates and standard deviations, similar in appearance to what would result from a logistic regression analysis which did not use multiple imputation. Wald statistics were computed and significance was assessed using these combined estimates.

Results

Sample Demographics

Participants were 1,071 adolescents who quit high school at some point during their schooling. Because of budget constraints, the small town was eliminated from the follow- up sample. Of these participants, 204 (19%) were non-Latino white males, 163 (15%) were non Latino white females, 400 (37%) were Mexican American males, and 304 (28%) were Mexican American females. The urban location contributed 795 (74%) participants, while 276 (26%) were from the mid- sized location. The age at dropout of these participants ranged from 13 to 21, with 6 participants (1%) having dropped out in 7th grade, 24 (2%) in 8th grade, 251 (23%) in 9th grade, 314 (29%) in 10th grade, 299 (28%) in 11th grade, and 177 (17%) in 12th grade. Note that a full 26% of the participants in the present study dropped out at 9th grade or earlier, a group previously not included in studies of returning dropouts.

Follow-up surveys were completed by 519 (49%) of the participants. Of these, 508 (47%) responded to the items regarding high school completion. There were 217 (43%) with no high school credentials, 175 (34 %) with GED certificates, and 116 (23%) with a high school diploma. Table 1 gives breakdowns of degree attainment for ethnicity and gender.

Table 1
Description of Degree Attainment, for Ethnicity and Gender

	No Degree	GED	Diploma
Male	114 (43%)	97 (36%)	56 (21%)
Female	103 (43%)	78 (32%)	60 (25%)
Non-Latino White	55 (34%)	34 (42%)	39 (24%)
Mexican American	162 (47%)	107 (31%)	77 (22%)

Table 2 gives means and standard deviations for the other variables considered in this study. The categorical variable (children) is included with a percent response to one category. The last column of Table 2 gives the percent of missing data for each independent variable considered in the present study. Possession of high school credentials was the only variable from the second wave of data utilized in this study. Accordingly, this variable has the greatest proportion of missing values. The variable measuring teacher caring was not included in the final two years of data collection, so it also has a high percentage of missing responses. Because of incomplete records, achievement tests were not always available for these students, resulting in the high percentage of missing data for this variable. Finally, since the socio-economic status measure included questions about both parents, many students who did not have two parents left blank the item inquiring about the absent parent. Multiple imputation was used to account for missing data in these and other variables.

Table 2
Description of Independent Variables

Continuous Variables				
Factor	Mean	Standard Deviation	Valid N	Percent Missing
Grade at dropout	10.31	1.10	1071	0.0%
Age at dropout	16.61	1.24	1061	0.9%
GPA	1.21	0.82	1023	4.5%
SES	0.05	0.67	844	21.2%

Test scores	54.15	23.45	807	24.6%
Teacher caring	2.71	1.04	790	26.2%
Categorical Variable				
Factor	Percent "yes"	Valid N	Percent Missing	
Have or expecting children	18.0%	1027	4.1%	

Table 3 gives means or percentages for each variable used in the logistic regression models, broken down by respondents and non-respondents (participants with and without follow-up data). Using statistical significance as a guide (alpha =.10), Mexican American participants and female participants were overrepresented in the follow-up sample. Mexican American participants comprised 68.6% of the respondents, as opposed to 63.0% of the nonrespondents, and 47.0% of the respondents were female, as opposed to 40.4% of the nonrespondents. Respondents scored slightly higher on achievement tests and were slightly younger.

Table 3
Means and Percentages, by Respondents and Non-respondents

Factor	**Respondent**	**Non-respondent**	**p**
Ethnicity	68.6% MA	63.0% MA	0.03
Gender	47.0% female	40.4% female	0.06
SES	0.04	0.07	0.48
Test scores	55.84	52.35	0.03
Age at dropout	16.54	16.67	0.09
Grade at dropout	10.30	10.32	0.79
GPA	1.23	1.20	0.54
Have or expecting children	18.2% yes	17.1% yes	0.64
Teacher caring	2.70	2.73	0.69

Distribution of Degree Attainment

Combining estimates of degree attainment across the twenty imputed data sets estimated that 40.8% of high school dropouts had no degree, 35.0% had a GED certificate, and 24.2% had a high school diploma.

Final Logistic Regression Models

Since the variables of interest were dichotomous (degree/no and diploma/GED), logistic regression was an appropriate analysis. For each logistic regression analysis in this section, predicted odds ratios are presented, and each estimate of an odds ratio is accompanied by a 95% confidence interval.

All estimates were obtained using multiple imputation (see Method). Typically, no more than ten data sets are needed for multiple imputation. However, preliminary examination of results using 10 imputed data sets indicated a greater amount of imputed data was needed to ensure stability of the estimates and to guarantee that variability due to imputation would be properly estimated. This is analogous to the practice of drawing a large sample to ensure that results will properly infer to the population. Therefore, 20 imputed data sets were used.

Tables 4 and 5 give the estimated odds ratios with 95% confidence intervals for significant factors in each model. Estimates of odds ratios are given in terms of the increase in odds for one unit change of the independent variable.

Degree vs. no degree. As described in Table 4, socio-economic status, test scores and age at dropout were the only variables shown to be significantly related to returning for a degree. Socio-economic status was positively associated with degree attainment, with a one point increase on the SES scale associated with an increase in the odds of returning of 1.34. A participant's test scores were positively related to degree attainment. A one point increase in average test score increased the odds of gaining a high school degree by a factor of 1.02, while a 10 point increase in test scores increased the odds of gaining a high school degree by a factor of 1.21 ($1.21 = 1.02^{10}$). Participants who dropped out as older adolescents were more likely to gain some form of high school credentials. For every year of age, the odds a participant would return for a degree was increased by 1.28. Thus, a participant who dropped out at age 18 was 2.12 times more likely to get a degree than a participant who dropped out at age 15 ($2.12 = 1.28^{3}$).

Table 4
Final Model Describing Degree Attainment: Variables From Previous Dropout Literature

Factor	Odds Ratio	95% Conf. Interval (Lower Bound, Upper Bound)	ß	se(ß)	t	df	p
SES	1.34	1.01, 1.79	0.29	0.145	2.03	91	0.045
Test scores	1.02	1.01, 1.03	0.02	0.005	4.07	50	0.000
Age at							

dropout	1.28	1.12, 1.47	0.25	0.069	3.57	117	0.001
Intercept	0.01	0.00, 0.10	-4.68	1.199	-3.90	104	0.000

Note. Dependent variable is degree/no degree.

High school diploma vs. GED. As described in Table 5, socio-economic status, test scores, children and grade at dropout significantly predicted the choice between a diploma or GED. Socio-economic status was positively associated with GED attainment. A one-point increase in the SES score was associated with an increase of 1.47 in the odds of GED attainment (an increase of .68 in the odds of diploma attainment). Higher test scores were also associated with GED attainment. Similar to the previous model, a one point increase in test scores was associated with an increase in the odds of GED attainment by a factor of 1.02, (an increase of .98 in the odds of diploma attainment) while a 10-point increase raised these odds by a factor of about 1.21. Having or expecting a child at the time of dropout was also associated with GED attainment. Degree holders having or expecting children were 1.92 times as likely to have a GED than a diploma (.52 times as likely to have a diploma than a GED). The amount of school a participant completed was a strong predictor of the type of degree held. A participant was approximately twice as likely to have a diploma for each increase in grade at dropout. To illustrate, someone who dropped out in 11th grade was estimated to be 7.46 times more likely to have a diploma than someone who dropped out in 8th grade.

Table 5
Final Model Describing Choice of Degree: Variables From Previous Dropout Literature

Factor	**Odds Ratio**	**95% Conf. Interval (Lower Bound, Upper Bound)**	**ß**	**se(ß)**	**t**	**df**	**p**
SES	0.68	0.47, 0.99	-0.38	0.188	-2.01	93	0.047
Test scores	0.98	0.97, 0.99	-0.02	0.006	-3.11	68	0.003
Grade at dropout	1.95	1.52, 2.51	0.67	0.126	5.31	79	0.000
Children	0.52	0.28, 0.95	-0.65	0.305	-2.14	111	0.035
			-		-		

Intercept	0.00	0.00, 0.03	6.26	1.296	4.83	79	0.000

Note. Dependent variable is diploma/GED.
Note. Children is Y/N.

Discussion

The present study extended and clarified previous work regarding degree attainment in high school dropouts. Previous studies had provided information on returning dropouts, but had been unable to include students who dropped out before 10th grade and students who were unavailable for subsequent followup. The present study was able to estimate relationships within the entire dropout population by including students who dropped out before 10th grade, and by using multiple imputation to estimate effects of students not included in followup data collection. Also, although previous studies had identified factors significantly associated with returning, each study contained omissions of factors deemed important by other studies. The present study was able to consider a broader view of the dropout's situation by collecting factors found significant in other studies, thus answering questions regarding the significance of these factors in the presence of other important factors. Finally, the present study compared dropouts without degrees to those with either a diploma or GED, performed in return studies only by Kolstad and Kaufman (1989).

Two separate logistic regression models were estimated, one discerning between dropouts with some sort of degree and those with no degree, the other discerning between dropouts with diplomas and those with GEDs. Results indicated that dropouts of higher socio-economic status, higher achievement test scores and greater age at dropout were more likely to attain some sort of degree. Analyses further showed that dropouts of higher socio-economic status, with higher test scores, and who dropped out having or expecting children were more likely to have GED certificates than high school diplomas, while those who dropped out in later grades were more likely to have diplomas than GEDs. Commonly identified factors such as ethnicity and gender were not significantly associated with either dependent measure.

How Many Gain Degrees?

One of the most striking findings of the present study is perhaps the simplest, that an estimated 59.2% of the high school dropouts from this study have returned to gain either a high school diploma or GED certificate. This result supports the assertions of previous studies that dropping out does not represent the end of a student's education. Further, it gives evidence of an increasing trend in degree attainment over the last ten years, as the estimate is 15.2% higher than the 44%

estimate given by Kolstad and Kaufman (1989). The difference is even more noteworthy when one considers that the present study includes participants who dropped out between seventh and twelfth grades, while the Kolstad and Kaufman study only included participants who dropped out in the tenth through twelfth grades. Grade has been shown in both studies to be positively associated with degree attainment, so the Kolstad and Kaufman estimates should be biased upward.

Also important to note from this finding is the role played by multiple imputation in reducing the bias introduced by participants who did not respond to the second wave of data collection. It has been commonly assumed (e.g., Kolstad, 1988) that dropouts who did not respond to subsequent waves of longitudinal data were "hard core" dropouts who were less likely to hold high school credentials. Such assumptions are admittedly conjecture, since degree estimates for this population were unavailable. The present study, however, estimated that dropouts who do not participate in subsequent data collection actually are slightly more likely to have some form of high school credential. Degrees were held by 57.2% of the participants who participated in the follow-up wave; estimates using multiple imputation indicated that 59.2% of the total sample holds high school credentials.

Degree vs. no degree

The results from this study indicate that generally, dropouts who gain some form of high school degree are of higher socio-economic status (SES), possess higher school capability (as measured by test scores), and are older when they drop out. The age and capability findings are consistent with previous literature and the fact that the present study proves these findings while accounting for earlier dropouts, participant nonresponse, and a wider breadth of factors suggests that these factors are robust. The SES finding clarifies some confusion in previous literature as to the significance of this factor. These findings stress the importance of targeting students of low SES and low capability, in addition to continued emphasis on early dropout prevention.

Possibly the greatest contribution of the model describing degree attainment is in the clarification of factors which are not significantly associated with returning for a degree. For instance, previous research had identified interactions involving ethnicity and SES, test scores and SES, gender and ethnicity, and gender and grade at dropout, but these interactions were not presented controlling for other important variables (Kolstad and Kaufman (1989); Kolstad and Owings (1986)). Results from the present study indicate that although significantly associated with degree attainment, grade at dropout and SES operate independently of other factors. Further, ethnicity and gender are not significant at all when controlling for other factors.

The fact that ethnicity was not found to be significant in these models should not be construed as a statement that ethnicity is unrelated to degree attainment. The univariate relationship between degree attainment and ethnicity indicated that non-Latino white dropouts are 1.73 times more likely to return to earn some form of high school degree (95% CI: 1.23, 2.43). However, the multivariate model indicated that SES, achievement test scores and age at dropout sufficiently explain the ethnic differences involved in the univariate effect. Further inspection of these results reveals that Mexican American dropouts display more risk in these factors than do non-Latino white dropouts. The test scores of Mexican American dropouts were on average 15.56 percentile points less than non-Latino white dropouts (95% CI: 12.08, 19.03), Mexican American dropouts were 2 months younger than non-Latino white dropouts (95% CI: .08, 3.85), and Mexican American dropouts averaged .56 of a standard deviation less on the SES scale than non-Latino white dropouts (95% CI: .48, .64). That these factors account for the univariate effect helps clarify some contradictory findings from previous literature on returning dropouts – if a study includes sufficient covariates, ethnic effects should be rendered insignificant.

Diploma vs. GED

Dropouts who chose a GED over a high school diploma were typically of higher socio-economic status (SES), possessed greater levels of school capability and were more likely to have children. Dropouts who chose to get a diploma rather than a GED typically dropped out at a later grade.

The grade in which a student drops out of high school is a strong predictor of which degree (s)he will attain. This is not unexpected – for a student who dropped out early in her/his high school career, finishing a high school diploma takes more time and effort than would attaining a GED. The magnitude of the grade/attainment relationship is large, more so than found by Kolstad and Kaufman (1989). This is likely due to the inclusion of younger dropouts in the present study.

Students of higher SES and of higher school capability were more likely to get a GED than a high school diploma. These results suggest that many students have the resources and capability needed to complete high school, but for some reason, school does not provide them with the fit they are looking for. It is possible that these students have specific aims in dropping out – given their higher social standing and ability, these students may have access to better jobs, schooling or training that require quick attainment of a high school credential. Or, these students may not have a specific goal in mind, but feel they have the ability to succeed at something, and that school does not afford them the environment to succeed as they want to. Also, it is possible these students dropped out with no future plans, then as they decided to return, they had better access to GED programs, GED information, etc., and just saw a GED as a quicker and easier way to get a degree.

Kolstad and Kaufman (1989) showed that participants who were parents were more likely to return for some kind of degree, while

Kolstad (1988) showed these students more likely to stay out of school. Results from the present study indicate that children don't affect overall degree attainment, but for those students who did attain a degree, those who had or were expecting children at the time of dropout were more likely to get a GED than a diploma. This is a reasonable finding, as many of these students would not be able to put forth the time required to finish a high school diploma. Also interesting is that there is no interaction with this factor and gender, indicating that the effect is the same for males as it is for females. Many studies (e.g., Rumberger, 1983; Wehlage & Rutter, 1986) suggest that females are more likely to drop out for child-related reasons. However, the return process is not that way.

Implications

The results and conclusions presented here have implications for education, and more specifically, dropout prevention and retrieval. Because of the breadth of factors considered, and the consideration of dropouts previously left out, this study has been able to clarify questions arising from previous research. In doing so, the present study has identified a group of factors which together appear to be most proximal in effecting degree attainment.

This study has joined previous research in affirming that dropping out is not the end of a student's education. Degree attainment in dropouts is a common occurrence, and results from the present study suggest it is more common now than ever. Despite these findings, the research devoted to dropping out of high school continues to weigh heavily toward studying causes and correlates of dropping out. It is imperative that research institutions and school systems greatly increase efforts to help dropouts return for degrees if in fact, they do drop out. In some schools, this may be an untapped resource in the fight to reduce dropout rates.

The simplicity of the final models should be helpful for practitioners. Based on factors considered here, degree attainment, whether by way of diploma or GED, can be explained in terms of a few important factors. Further, the decision to return for a degree operates similarly regardless of gender or ethnicity (Mexican American or non- Latino white). Therefore, the models estimated here suggest that dropout retrieval programs (and some facets of dropout prevention programs) can possibly be simplified, streamlined, and ultimately, less expensive.

Important for practitioners also is the finding that of dropouts who return for degrees, GED-holders on average have higher school capability. As described above, the reasons why these factors are significant are not evident. It is clear that these students have capability to do school work, and seemingly, school is not a good fit for them. However, it seems that these students are walking a dangerous line in opting for a GED instead of a diploma, since attainment of a high school diploma is associated with more labor and economic success than is attainment of a GED certificate (Cameron & Heckman, 1993; Passmore, 1987). This is not to say that for all students, a high school degree is a better choice than a GED, but research suggests that a high school degree is better for most students unless there is a demonstrated situation where the GED would be better. Therefore, schools should persevere to provide opportunities which could channel these students toward diploma attainment, an endeavor which will likely be more positive for the student in the long run.

Although there are positives associated with the simplicity of these models, the specific factors identified are also discouraging for practitioners attempting to change the life trajectories of these dropouts. Starkly obvious from the models presented here is the fact that degree attainment in dropouts is a function of factors in a student's life which are very difficult for schools to change. Despite the fact that this study has clarified many issues regarding returning dropouts, it is now clear that different frameworks should be explored in order to identify factors which are more easily changed by practitioners.

Educational research can inform decisions on where to turn next. Finn and Rock (1997) have argued that the research on academic success has placed undue focus on relatively constant characteristics of the individual, and that more focus should be placed on factors which can be changed by educators. Augmenting this notion is the assertion by Alva (1991), that subjective student appraisals are very important in the evaluation of the student's educational experience. School structure could play a role in helping dropouts return, in fact, many researchers (e.g., Finn & Rock, 1997; Wehlage & Rutter, 1986) believe that the secret to educating at-risk students lies in the alteration of factors related to school. Judicious alteration of school factors could serve to aid in positive alteration of individual factors.

Thus, there is room for future research on returning dropouts to expand into a less restrictive framework. Attention should be turned to more positive correlates, ones associated with academic success rather than failure, aiming to identify areas where both the school and student can more easily exact positive change. Candidates for such expansion include the roles of attitudinal factors, which are more malleable and more internal to the student, factors pertaining to peers and family, factors pertaining to schools, such as teacher attitudes and communication, and school opportunities and definitions of success.

Conclusion

The present study has extended previous research on dropouts who gain degrees. This study has found, as have other studies, that high school dropouts frequently return to gain degrees of some form, a finding which underscores the need for more research in this area. This study has also provided clarification of correlates of degree attainment. In doing so, it has presented a neat, concise package of factors which influence returning for a degree. Although concise, this group of factors also presents a problem, in that they are factors which are difficult to change in order to create a more positive situation for a dropout. Hence, this study has illuminated the need for additional studies on returning dropouts which can build upon knowledge presented here. Such studies should endeavor to consider more positive correlates of returning, ones which can more easily be effected by schools and practitioners.

Notes

1. This study was supported by the National Institute on Drug Abuse under grant R01 DA 04777.
2. The author is especially grateful to Randall C. Swaim, and also to Brian Cobb, Bill Timpson and Cori Mantle-Bromley for their insightful comments. In addition, thanks to Ernest Chavez for making available the data used in this study.

References

Alva, S.A. (1991). Academic invulnerability among Mexican-American students: The importance of protective resources and appraisals. *Hispanic Journal of Behavioral Sciences,13* (1), 18-34.

Borus, M.E. & Carpenter, S.A. (1983). A note on the return of dropouts to high school. *Youth and Society, 14,* 501-507.

Cameron, S.V., & Heckman, J.J. (1993). The nonequivalence of high school equivalents. *Journal of Labor Economics, 11* (1), 1-47.

Catterall, J.S. (1987). On the social costs of dropping out of school. *High School Journal, 71,* 19-30.

Chavez, E.L., Deffenbacher, J.L, & Wayman, J.C. (1996). A longitudinal study of drug involvement in Mexican American and white non-Hispanic high school dropouts, academically at risk students and control students. *Free Inquiry in Creative Sociology, 24,* 185-193.

Chavez, E.L., Oetting, E.R., & Swaim, R.C. (1994). Dropout and delinquency: Mexican-American and Caucasian non- Hispanic youth. *Journal of Clinical Child Psychology, 23* (1), 47-55.

Chuang, H. L. (1997). High school youths' dropout and re-enrollment behavior. *Economics of Education Review, 16* (2), 171-186.

Ekstrom, R.B., Goertz, M.E., Pollack, J.M., & Rock, D.A. (1986). Who drops out and why? Findings from a national study. *Teachers College Record, 87,* 356- 373.

Elliott, D.S., Huizinga, D., & Ageton, S. S. (1985). *Explaining* *delinquency and drug use.* Newbury Park, CA: SAGE.

Finn, J.D., & Rock, D.A. (1997). Academic success among students at risk for school failure. *Journal of Applied Psychology, 82* (2), 221-234.

Graham, J.W., Hofer, S.M., Donaldson, S.I., MacKinnon, D.P., & Schafer, J.L. (1997). Analysis with missing data in prevention research. In K. Bryant, M. Windle, & S. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research.* (pp. 325-366). Washington, D.C.: American Psychological Association.

Kaufman, P. (1988). *High school dropouts who return to school.* Unpublished doctoral dissertation, Claremont Graduate School, Claremont, CA.

Kolstad, A.J. & Owings, J.A. (1986). *High school dropouts who change their minds about school.* Washington, DC: Office of Educational Research and Improvement. (ERIC Document Reproduction Service No. ED 275 800)

Kolstad, A.J. & Kaufman, P. (1989, March). *Dropouts who complete high school with a diploma or GED.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Morrow, G. (1986). Standardizing practice in the analysis of school dropouts. *Teachers College Record, 87*, 342-355.

Passmore, D.L. (1987). *Employment of young GED recipients.* American Council on Education Research Brief (14).

Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys.* New York: Wiley.

Rumberger, R.W. (1983). Dropping out of high school: The influence of race, sex and family background. *American Educational Research Journal, 20,* 199- 220.

Rumberger, R.W. (1987). High school dropouts: A review of issues and evidence. *Review of Educational Research, 57* (2), 101-121.

Schafer, J.L. (1999) NORM: Multiple imputation of incomplete multivariate data under a normal model, version 2. Software for Windows 95/98/NT, available from http://www.stat.psu.edu/~jls/misoftwa.html.

Schafer, J.L. (1997). *Analysis of incomplete multivariate data.* New

York: Chapman and Hall.

Waxman, H.C., Huang, S.L., Padron, Y.N. (1997). Motivation and learning environment differences between resilient and nonresilient Latino middle school students. *Hispanic Journal of Behavioral Sciences, 19* (2), 137- 155.

Wehlage, G.G., & Rutter, R.A. (1986). Dropping out: How much do schools contribute to the problem? *Teachers College Record, 87*, 374-392.

About the Author

Jeffrey C. Wayman, Ph.D.
Tri-Ethnic Center for Prevention Research
101 Sage Hall, Colorado State University
Fort Collins, CO 80523.
Phone: (970) 491-6969
Fax: (970) 491-0527.

Email: wayman@lamar.colostate.edu

Jeff Wayman is a Research Associate with the Tri-Ethnic Center for Prevention Research, at Colorado State University. He holds a Ph.D. in Education and a Masters in Statistics. His current educational research interests include at-risk students, educational resilience, and cultural issues, and how these issues relate to teacher training and school reform. Current methodological interests include missing data issues and multilevel modeling.

Copyright 2001 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* **is epaa.asu.edu**

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple University of Wisconsin	Greg Camilli Rutgers University
John Covaleskie Northern Michigan University	Alan Davis University of Colorado, Denver
Sherman Dorn University of South Florida	Mark E. Fetler California Commission on Teacher Credentialing
Richard Garlikov hmwkhelp@scott.net	Thomas F. Green Syracuse University
Alison I. Griffith York University	Arlen Gullickson Western Michigan University
Ernest R. House University of Colorado	Aimee Howley Ohio University
Craig B. Howley Appalachia Educational Laboratory	William Hunter University of Calgary
Daniel Kallós Umeå University	Benjamin Levin University of Manitoba
Thomas Mauhs-Pugh Green Mountain College	Dewayne Matthews Western Interstate Commission for Higher Education
William McInerney Purdue University	Mary McKeown-Moak MGT of America (Austin, TX)
Les McLean University of Toronto	Susan Bobbitt Nolen University of Washington
Anne L. Pemberton apembert@pen.k12.va.us	Hugh G. Petrie SUNY Buffalo
Richard C. Richardson New York University	Anthony G. Rud Jr. Purdue University
Dennis Sayers Ann Leavenworth Center for Accelerated Learning	Jay D. Scribner University of Texas at Austin
Michael Scriven scriven@aol.com	Robert E. Stake University of Illinois—UC
Robert Stonehill U.S. Department of Education	David D. Williams Brigham Young University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México) Universidad de Guadalajara adrianacosta@compuserve.com	J. Félix Angulo Rasco (Spain) Universidad de Cádiz felix.angulo@uca.es
Teresa Bracho (México) Centro de Investigación y Docencia Económica-CIDE	Alejandro Canales (México) Universidad Nacional Autónoma de México

bracho dis1.cide.mx

canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

other vols. | abstracts | editors | board | submit | comment | subscribe | search

This article has been retrieved 1609 times since February 14, 2001

other vols. | abstracts | editors | board | submit | comment | subscribe | search

Education Policy Analysis Archives

Volume 9 Number 5 **February 14, 2001** **ISSN 1068-2341**

A peer-reviewed scholarly journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2001, the **EDUCATION POLICY ANALYSIS ARCHIVES**.
Permission is hereby granted to copy any article
if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

How the Internet Will Help Large-Scale Assessment Reinvent Itself

Randy Elliot Bennett
Educational Testing Service
U.S.A.

Abstract
Large-scale assessment in the United States is undergoing enormous pressure to change. That pressure stems from many causes. Depending upon the type of test, the issues precipitating change include an outmoded cognitive-scientific basis for test design; a mismatch with curriculum; the differential performance of population groups; a lack of information to help individuals improve; and inefficiency. These issues provide a strong motivation to reconceptualize both the substance and the business of large-scale assessment. At the same time, advances in technology, measurement, and cognitive science are

> providing the means to make that reconceptualization a reality. The thesis of this paper is that the largest facilitating factor will be technological, in particular the Internet. In the same way that it is already helping to revolutionize commerce, education, and even social interaction, the Internet will help revolutionize the business and substance of large-scale assessment.

Whether for educational admissions, school and student accountability, or public policy, large-scale assessment in the United States is undergoing enormous pressure to change. This pressure is most evident with respect to high-stakes tests, like those used for grade promotion or college entrance. However, it is becoming apparent for lower-stakes survey instruments too, like the National Assessment of Educational Progress (NAEP) (e.g., Pellegrino, Jones, & Mitchell, 1999).

Several factors underlie the pressure to change. First, whereas our tests have incorporated many psychometric advances, they have remained separated from equally important advances in cognitive science, in essence measuring the same things in ever more technically sophisticated ways. Although decades of research have documented the importance of such cognitive constructs as knowledge organization, problem representation, mental models, and automaticity (Glaser, 1991), our tests typically do not account for them explicitly. As a result, our tests probably owe more to the behavioral psychology of the early 20th century than to the cognitive science of today (Shepard, 2000).

A second factor is the mismatch with the content and format of curriculum, a criticism more true of the developed ability tests commonly used in postsecondary admissions than of school achievement measures, but relevant to the latter too. The mismatch arises in part from the fact that the elemental, forced-choice problems dominating many tests are effective indicators of skills and abilities, and thus provide an efficient means for estimating student standing on those constructs. However, the mismatch becomes problematic because of the increasing attention being paid to test preparation. Although persistent direct training on these indicator tasks may increase test performance, it certainly is not the best way to improve construct standing. Further, it distracts attention from other, arguably more critical, learning activities (Frederiksen, 1984).

Differential performance of population groups is another factor. Because of the curricular mismatch, it is easy to blame group differences on purported bias in the test and more difficult to create a convincing defense than it would be if the tests were strongly linked to learning goals. In a high-stakes decision setting like admissions, tests become a lightning rod for the failure of schools and society to educate all groups effectively. With the potential elimination of affirmative action in university admissions, there is no politically acceptable choice but to reduce the role of such tests. California, Texas, Florida, and Pennsylvania are proposing to admit, or have begun admitting, all students with high-school rank above a certain point to their state higher education systems. At the same time, promotion tests tied to state curricular standards are being put into place to encourage schools to teach all students valued skills. Although in Texas one such test was challenged in court on the basis of differential performance, that challenge was rejected (Schmidt, 2000). This rejection suggests that when well- constructed tests closely reflect the curriculum, group differences should become more an issue of instructional inadequacy than test inaccuracy (Bennett, 1998).

As attention shifts to the adequacy of instruction, the ability to derive meaningful information from test performance becomes more critical. A weak connection between test and curriculum insures that the value of feedback for the examinee will be limited. Even for tests where the connection is stronger, feedback is still too often of marginal value, in part because of the additional cost and processing time that would be incurred. For achievement surveys like NAEP, which offer no information to individuals, schools, or districts, motivation to participate is undoubtedly diminished.

Finally, there is efficiency. Testing programs are expensive to operate. That expense gets passed on to taxpayers for a state or federal test like NAEP, or directly to examinees in the case of admissions measures. Further, to be maximally useful, test results are needed quickly. Rapid information delivery is certainly a requirement in the education policy arena, where the results of national surveys may sometimes take years to produce. It is also increasingly true in the admissions context, where more rapid feedback is needed not only for early decisions, financial aid, and the rolling acceptances that are beginning to characterize some distance learning programs, but also for guidance and placement.

Will reinvention solve all of these problems? Of course not. But I do believe it will allow us to make significant progress on each of them.

Does reinvention mean abandoning educational testing as it now exists? No. It only means combining the best of the old with the most promising of the new to engineer radical improvements.

The Promise of New Technology

Radical improvements in assessment will derive from advances in three areas: technology, measurement, and cognitive science (Bennett, 1999). Of the three, new technology will be the most influential in the short term and, for that reason, I focus on it in this paper. New technology will have the greatest influence because it—not measurement and not cognitive science—is pervading our society. Billions of dollars are being invested annually to create and make commonplace powerful, general technologies for commerce, communications, entertainment, and education. Due to their generality, these technologies can also be used to improve assessment.

These technological advancements revolve primarily around the

Internet. The Internet is (or will be) interactive, broadband, switched, networked, and standards-based. What does that mean?

- *Interactive* means that we can present a task to a student and quickly respond to that student's actions.
- *Switched* means that we can engage in different interactions with different students simultaneously. In combination, these two characteristics (interactive and switched) make for individualized assessments.
- *Broadband* means that those interactions can contain lots of information. For assessment tasks, that information could include audio, video, and animation. Those features might make tasks more authentic and more engaging, as well as allow us to assess skills that cannot be measured in paper and pencil (Bennett, Goodman, Hessinger, Ligget, Marshall, Kahn, & Zack, 1999). We might also use audio and video to capture answers, for example, giving examinees choice in their response modalities (typing, speaking, or, for a deaf student, American Sign Language).
- *Networked* indicates that everything is linked. This linkage means that testing agencies, schools, parents, government officials, item writers, test reviewers, human scorers, and students are tied together electronically. That electronic connection can allow for enormous efficiencies.
- Finally, *standards-based* means that the network runs according to a set of conventional rules that all participants follow. That fact permits both the easy interchange of data and access from a wide variety of computing platforms, as long as the software running on those platforms (e.g., Internet browsers), adheres to those rules too. (Note 1)

As an embodiment of these characteristics, what does the Internet afford? It affords the potential to deliver efficiently on a mass scale individualized, highly engaging content to almost any desktop; get data back immediately; process it; and make information available anywhere in the world, anytime day or night. Paper delivery cannot compete with this potential.

The Internet is, of course, not being built to service the needs of large-scale assessment. It is, instead, being built for e-commerce: to sell products and services over the web to consumers and to businesses directly. Coincidentally, the capabilities needed for e-commerce are essentially those needed for e-assessment:

- interactive (so that products can be offered and orders transacted),
- switched (so different business transactions can be conducted with different customers simultaneously),
- broadband (so that those offers can be as engaging and enticing as possible),
- networked (so that product offers, orders, shipping, inventory,

and accounting can be integrated), and
- standards-based (so that everyone can get to it, regardless of computing platform).

Will we be able to count on continued investment in the Internet to support its use as a delivery medium? By any measure, the Internet and use of it, has grown dramatically, to say the least. As a communications medium, the Internet last year surpassed the telephone, with 3 billion email messages sent each day (Church, 1999). The number of unique URLs (web-page directory and subdirectory addresses) has grown from just under a billion in 1998 to a projected 3 billion in 2000 ("Big fish," 1999). In the United States, the percentage of homes with Internet access has increased from 26% in December 1998 to 42% in August 2000 (U.S. Department of Commerce, 2000). (Note 2) Worldwide, the number of users has grown from somewhere between 117 to 142 million in 1998 to about 400 million in 2000 ("Big fish," 1999; Global Reach, 2000; "How many online?", 2000). Finally, the number of host computers has gone from about 30 million to 75 million from January 1998 to January 2000 ("Internet domain survey host count," 2000). This phenomenal growth may slow as investment subsides from the speculative rates of the past few years. However, the vast size of the Internet and its user base constitute a critical mass that should continue to attract substantial capital.

For commerce, the promise of the Internet is all about being faster, cheaper, and better. Two "laws" of the digital era illustrate this promise. Moore's Law predicts the doubling of computational capability (specifically, at the level of the microchip) every 18 months. As Negroponte (1995) has explained, what filled a room yesterday is on your desk today and will be on your wrist tomorrow. Metcalfe's Law says that the value of a network increases by the square of the number of people on it. The true value of a network is, thus, less about information and more about community (Negroponte, 1995). One can see this effect clearly in eBay, the online auction broker (Cohen, 1999). Each new user potentially benefits every other existing user because every eBay member can be both buyer and seller. (Note 3) Metcalfe's law is playing out well beyond eBay. Online business-to-business auction brokers are appearing in a variety of industries, including natural gas, electricity, steel, and bandwidth (Friedman, 2000, pp. 386-387; Gibney, 2000).

Another illustration of this cheaper-faster-better result is the effect of the Internet on the traditional relationship between *richness* and *reach*, where *richness* is the depth of the interaction that a business can have with a customer and *reach* is the number of customers that a business can contact through a given channel. Traditionally, one limited the other. That is, a business could attain maximal reach but only limited richness. For example, through direct mail, broadcast, or newspaper ads a company could communicate with many people but have a meaningful interaction with none of them. Similarly, a business could attain maximal richness but limited reach.

Via personal contact (e.g., door-to-door sales), very deep interactions can occur, but with only a relatively small number of people. What has the Internet done? It has transformed the relationship between richness and reach by allowing businesses to touch many people in a personalized but inexpensive way (Evans & Wurster, 2000). What does richness *with* reach make for? It makes for *mass customization*.

We can already see the effects in Dell Computer Corporation's business model. Customers can log onto Dell's Internet site (www.dell.com), choose from a menu of basic machine designs, and then configure a particular design to meet their needs. A second example is Radio.SonicNet (http://radio.sonicnet.com/splash.asp). Radio.SonicNet allows one to pick from a variety of music styles, choose artists within that style, and indicate how frequently each artist should play. The end result is a radio station uniquely tuned to the individual and continually interesting; it always plays what you like but you never know exactly what it is going to play. As a final example, consider Customatix (www.customatix.com/customatix/common/homepage/HomepageGeneral.po), which allows you to design your own shoes using up to *three billion trillion* combinations of colors, graphics, logos and materials per shoe. You design them. They build them. And *nobody else* is likely to have exactly the same ones.

Reinventing Assessment

Reinventing the Business

There are two major dimensions to reinventing assessment. One is the business of assessment. This dimension centers on the core processes that define an enterprise. In many cases, those core processes can become many times more efficient because moving bits is faster and easier than moving atoms (Negroponte, 1995); that is, electronically processing information is far more cost effective than physically manipulating things.

For large-scale testing programs, some examples of the potential for electronic processing are in:

- developing tests, making the items easier to review, revise, and automatically morph into still more items (e.g., Singley & Bennett, in press) because the items themselves are digitally represented;
- delivering tests, eliminating the costs of printing, warehousing, and shipping tons of paper;
- presenting dynamic stimuli like audio, video, and animation, making the need for specialized testing equipment (e.g., audio cassette recorders, VCRs) obsolete (Bennett, Goodman, Hessinger, Ligget, Marshall, Kahn, & Zack, 1999);
- transmitting some types of complex constructed responses to human graders, removing the need to transport, house, and feed the graders (Odendahl, 1999; Whalen & Bejar, 1998);

- scoring other complex constructed responses automatically, reducing the need for human reading (Burstein et al., 1998; Clauser et al., 1997); and
- distributing test results, cutting the costs of printing and mailing reports.

To get a sense of how reinventing the business of assessment might affect testing organizations, take a look at reference book publishing, in particular the case of *Encyclopaedia Britannica* (Evans & Wurster, 2000; Landler, 1995; Melcher, 1997). *Encyclopaedia Britannica* was established in Scotland in 1768. It is the oldest and most famous encyclopedia in the English-speaking world. By 1990, its sales had reached $650 million per annum. But then suddenly, *Britannica's* fortunes drastically changed. In 1996, the company was sold for less than half its net worth (i.e., the value of its assets, including its encyclopedia inventory, minus its liabilities). That same year, it eliminated its entire door-to-door North American sales force. By 1998, sales had fallen 80%. What happened?

What happened was that the reference book business was reinvented because of the emergence of new technology. At its peak, *Britannica* was a 32-volume set of books costing well over $1,000. In 1993, Microsoft introduced *Encarta* on CD-ROM for under $100 and even though *Britannica* was much more comprehensive, the difference for most people wasn't worth an extra $900+. Initially, *Britannica* did not respond as it didn't take the threat from *Encarta* seriously. But when it did respond, it did so ineffectively because *Britannica* wouldn't fit on a single CD-ROM and because the company's large sales force wasn't suited to selling software. But, ultimately, *Britannica* wasn't ready to cannibalize its existing paper business to enter this new electronic one.

Why is this story important? It's important because similar (though less extreme) scenarios are playing themselves out now in individual investing, book selling, travel planning, music distribution, long distance telephony, and even business-to-business transactions. (As to the last, Cisco Systems makes 90% of its revenue from business-to-business transactions done over the Internet [Cisco Systems, Inc., 2000]). These reinvention scenarios are forcing organizations—including some in educational assessment—to come quickly to grips with where new technology *will* and *will not* help core business processes.

As should be obvious, technology-driven changes in business processes can occur quickly and their consequences can be significant for the organizations that service a particular market. In fact, if radical and pervasive enough, process changes can force shifts in the substance of the business itself. So, although reinventing the business of assessment by incorporating technology into specific assessment processes is about trying to achieve the efficiencies needed to remain competitive today, reinventing the substance of assessment—most fundamentally, the reason we do it—is not about today. It's about tomorrow.

Reinventing the Substance

The populations seeking education are changing and so are their purposes for learning. At the college level, just *16%* of students fit the traditional profile: 18-22 years old, full-time, on-campus resident (Levine, 2000a). This is not because fewer 18-22 year olds are going to college. It is because more adults are. The adult cohort is, in fact, the fastest growing segment in postsecondary education (Kerrey & Isakson, 2000). Working adults over age 24 constitute some 44% of college students ("Education prognosis 1999," 1999).

Why are so many adults returning to college? Over the past 25 years, employer demand in the U.S. has shifted toward higher educational qualifications, as indicated by an increasing premium paid for those with a college degree (Barton, 1999). But in addition to this rise in entry qualifications, the knowledge required to maintain a job in many occupations is changing so fast that 50% of all employees' skills are estimated to become outdated within 3-5 years (Moe & Blodget, 2000). Witness any job that requires interaction with information technology (IT), which is a growing proportion of jobs. In fact, by 2006 almost half of all workers will be employed by industries that are either major producers or intensive users of IT products and services (Henry et al., 1999).

So, more people want postsecondary education because they need to have it if they want to become—and stay—employed. And, more of these individuals are nontraditional students who may work, travel in their jobs, or have families. For these people, physically attending classes is not always feasible, let alone convenient. (Note 4)

This population's unmet educational need is increasingly becoming the target of distance learning. According to the National Center for Education Statistics, between fall 1995 and 1997-98, the percentage of higher education institutions offering distance learning courses increased by one-third (from 33% to 44%), and the number of course offerings and enrollments approximately doubled (Lewis et al., 1999). But although many institutions have delivered distance learning via mail, radio, or television for years, this growth is not in those media. Rather, it is distance learning via the Internet that is booming. Among all higher-education institutions offering any distance learning, the percentage of institutions using asynchronous Internet-based technologies nearly tripled, from 22% in 1995 to 60% in 1997-1998. More recent data from Market Data Retrieval (MDR) confirm the trend ("Report: College Net use growing," 2000). MDR relates that, as of the 1999-2000 academic year, 34% of two- and four-year colleges offered accredited degree programs via computer, up from 15% the year before. As of 2000, U.S. institutions reportedly offered more than 6,000 accredited courses on the Web and, by 2002, over 2 million students will be enrolled, a tripling of the 1998 enrollment (Moe & Blodget, 2000).

At the same time, Internet-based distance learning is finding its way into high school. The need is generated by home-schooled students (of which there are over 1 million in the US), districts without a full complement of qualified teachers, and the children of migrant workers. So-called "virtual high schools" have emerged in Alabama, Arizona, California, Florida, Illinois, Indiana, Kentucky, Maryland, Massachusetts, Michigan, Missouri, Nebraska, New Mexico, and Utah (Carr, 1999; Carr & Young, 1999; Kerrey & Isakson, 2000). These programs can cross state lines, with offerings open to students regardless of residence. Of particular note is that both the University of Missouri at Columbia High School and the Indiana University High School have been granted accreditation by the North Central Association of Colleges and Schools (Carr, 1999). Accreditation means that students can apply course grades earned through these online institutions toward their high-school graduation. Both programs offer more than 100 high school courses.

The growth of Internet-based distance learning will have a significant impact upon traditional education. For one, it may threaten the existence of established institutions (Dunn, 2000; Levine, 2000b). Many in the private sector see education as a huge industry that produces mediocre results for a high cost. If the private sector can leverage new technologies, like distance learning, to deliver greater value, the institutions that dominate education today will not be the leaders tomorrow. The rapid growth of for-profit education companies (e.g., the University of Phoenix), and the seemingly endless creation of well-capitalized new ones (e.g., UNext, Caliber, KaplanCollege.com, University Access, K12), suggests that a serious challenge to the existing order is well underway. The gravity of the threat is evident in how non-profits have responded. Cornell University, Columbia University, the University of Maryland, and New York University, among others, have each announced their own for-profit distance learning subsidiaries (Carr, 2000a)!

A second reason that the growth of Internet-based distance learning will influence traditional education is that regardless of its impact on nonprofit institutions, the distance learning industry will produce sophisticated software that everyone can use, in school and out. Both Dunn (2000) and Tulloch (2000) suggest that this occurrence will blur the distinctions between distance learning and local education. APEX offers an example (http://apex.netu.com/). This company markets online Advanced Placement (AP) courses, targeting districts that want to offer AP but which do not have qualified teachers. Districts can, thus, use APEX offerings on site. (Note 5)

The considerable potential of online learning—local or distance—is reflected in a report to the President and Congress of the bipartisan Web-Based Education Commission (Kerrey & Isakson, 2000). The Commission reached the following conclusion:

> The question is no longer *if* the Internet can be used to transform learning in new and powerful ways. The

> Commission has found that it can. Nor is the question *should* we invest the time, the energy, and the money necessary to fulfill its promise in defining and shaping new learning opportunity. The Commission believes that we should. (p. 134, italics in original)

If acted on, the consequences of this statement for assessment are profound. As online learning becomes more widespread, the substance and format of assessment will need to keep pace. Another quote from the Commission's report:

> Perhaps the greatest barrier to innovative teaching is assessment that measures yesterday's learning goals...Too often today's tests measure yesterday's skills with yesterday's testing technologies—paper and pencil. (p. 59)

So, as students do more and more of their learning using technology tools, asking them to express that learning in a medium different from the one they typically work in will become increasingly untenable, especially where working with the medium is part of the skill being tested (or otherwise impacts it in important ways). Searching for information using the World Wide Web or writing on computer are examples. (Note 6)

These changes in learning methodology offer exciting possibilities for assessment innovation. On site or off, an obvious result of delivering courses via the Internet is the potential for embedding assessment, perhaps almost seamlessly, in instruction (Bennett, 1998). Since students respond to instructional exercises electronically, their responses can be recorded, leaving a continuous learning trace. Depending upon how the course and the assessment are designed, this information could conceivably support a sophisticated model of student proficiencies (Gitomer, Mislevy, & Steinberg, 1995). That model might be useful both for dynamically deciding what instruction to present next and for making more global judgments about what the student knows and can do at any given point.

In addition to assessment embedded in Internet-delivered courses, one can imagine Internet-delivered-assessment embedded in traditional classroom activity. Such assessment might take the form of periodically delivered exercises that both teach and test. In this scenario, the exercises would be standardized and performance might serve, depending upon the level of aggregation, to indicate individual, classroom, school, district, state, or national achievement. Thus, these exercises could serve summative as well as formative purposes and be useful to individuals as well as institutions. If the exercises were of high enough quality, such a model might improve the motivation to participate in voluntary surveys like NAEP.

There are, to be sure, many difficult issues:

1. How can we generate comparable inferences across students and

institutions when variation in school equipment may cause items to display differently from one student to the next, potentially affecting performance?

2. How can we deliver assessment dependably given the unreliable nature of computers and the Internet, and the limited technical support available in most schools?
3. How might we make sense of the huge corpus of data that the electronic recording of student actions might provide?
4. How would student learning be affected by knowing that one's actions are being recorded?
5. How can we prevent assessments that serve both instructional and accountability purposes from being corrupted by unscrupulous students or school staff?
6. How can we manage the costs of online assessment?
7. How can we assure that all parties can participate?

Let's, for the moment, turn to this last issue.

Are the Schools Ready?

A continuing concern with such reinvention visions is whether schools (and students) are ready technologically and, in particular, what to do about technology differences across social groups. The National Center for Education Statistics (NCES) reports that as of September 1999, 95% of schools were connected to the Internet, up from 35% in 1994 (NCES, 2000). Schools in *all* categories, (i.e., by grade level, poverty concentration, and metropolitan status), were equally likely to have Internet access. Further, most schools had dedicated lines: only 14% were using dial-up modem, a slower and less reliable access method.(Note 7)

Clearly many of these schools could have only a single connected machine and that machine could be the one sitting on the principal's desk. How many classrooms were actually wired? According to NCES (2000), as of September 1999, 63% of all instructional rooms had Internet access (up from 3% in 1994, a 20-fold increase in five years). The ratio of students to Internet-connected computers was 9:1, down from 12:1 only a year earlier. These are staggering numbers, for they imply that classrooms are connecting to the Internet at a very rapid rate.

This success is in no small part due to federal efforts. The government's *e-rate* program has been giving public schools and libraries discounts of up to 90% on phone service, Internet hook-ups, and wiring for several years ("FCC: E-rate subsidy funded," 2000). In total, the program has committed 3.65 billion dollars to over 50,000 institutions, helping connect more than one million public school classrooms (Kennard, 2000). In addition, 70% of the program's last round of funding went to schools in the lowest income areas.

However, even with these very significant efforts, there continue to be equity issues. As of September 1999, in high poverty schools, the ratio of students to Internet computers was 16 to 1. In low poverty schools, it was less than half that amount—7 to 1 (NCES, 2000).

What should we conclude? Certainly, with few exceptions, it would be impossible to deliver large-scale assessment via the Internet today. But the trend is clear: the infrastructure is quickly falling into place for Internet delivery of assessment to schools, perhaps first in survey programs like NAEP that require only a small participant sample from each school, but eventually for inclusive assessments delivered directly to the desktop. As evidence, witness the requests-for-proposals recently released by the state education departments of Oregon, Virginia, and Georgia for building Internet-delivered, state-assessment systems (Department of Education, 2000; Virginia Department of Education, undated, State of Georgia, 2001).

Assuming that every classroom is wired, will all students then have the technology skills needed to take tests on-line? Clearly, more students are becoming computer-familiar every day and developing such skills is a national educational technology goal (Riley, Holleman, & Roberts, 2000). But, as Negroponte (1995) suggests, computer familiarity is really the wrong issue. The secret to good interface design is to make it *go away*. Thus, advances in technology will eventually eliminate the need to be computer familiar. After *nomadic* computing, which we are now entering with the proliferation of wireless Internet devices and personal digital assistants, comes *ubiquitous* computing (Olsen, 2000)—the embedding of new technology into everyday items. Inventions like "radio" paper (Gershenfeld, 1999, p. 18; Maney, 2000; "NCS secures rights," 2000) may allow students to interact with computers in the same way that they interact with paper today. Smart desks are another likelihood, in which case a test may be electronically delivered, quite literally, to every desktop.

In the U. S., then, we may see a future in which every classroom is wired and every student can easily take tests on line. What of the rest of the world? To be sure, the Internet is an American phenomenon. It derives from research sponsored by the Defense Department in the 1960's (Cerf, 1993). As a result of this history, the overwhelming majority of users were, until very recently, from our shores. At this writing, over *60%* of Net users reside *outside* of the United States and the foreign growth rate now exceeds the domestic one ("How many online?", 2000; "U.S. dominance seen slipping," 2001).

The largest numbers of foreign Internet users are, of course, in developed nations. These nations have the telecommunications infrastructure and citizens with enough disposable income to afford the trappings of Internet use. But what about developing nations? Will they be left irretrievably behind? The challenges for these nations are undoubtedly great. Over time, however, we should see significant progress in building the infrastructure and the user base here too (Cairncross, 1997; Fernandez, 2000). This progress will occur for at least two reasons. First, the cost of technology has been dropping precipitously and, by Moore's law, will continue to decline. Further, because the future of computing is undoubtedly in wireless devices (Grice, 2000), a telecommunications infrastructure will be much cheaper to acquire than the land-lines of old. Second, as Metcalfe's law suggests, markets will become all the more valuable as they are interconnected. (Witness the global economy and the economic benefits resulting to nations from integration with it.) That developing nations join the e-commerce network means greater opportunity for all. It means more vendor choice for the people of developing nations; more opportunity for developed nations to serve these markets; and a new opportunity for third-world businesses themselves to compete globally. (Note 8)

The same holds true for assessment. The Internet will make it easier for developing nations to get access to assessment services from elsewhere and for those nations to distribute their own assessment services regionally or around the world. This ease of access and distribution should make it possible to form international consortia. Such consortia will be able to assemble technical resources that a single nation might not be able to acquire. In addition, those consortia may be able to purchase services from others more efficiently than nations could obtain individually. Finally, an electronic network should make it easier to participate in international studies, bringing the benefits of benchmarking to nations throughout the world.

But is Technology-Based Assessment Really Worth the Investment?

One of the largest instantiations of technology-based assessment to date is computer-based testing (CBT) in postsecondary admissions. As programs like the Graduate Record Examinations, the Graduate Management Admission Test, and the Test of English as a Foreign Language have found, CBT can be enormously costly. Being among the first large-scale programs to move to computer, they bore the brunt of creating the infrastructure for what was essentially a new business. The building of that infrastructure was initiated in the early 1990's *before* test developers knew how to create tests for computer, *before* computers were widely available for individuals to take tests on, and *before* the Internet was ready to bring those tests to students. In essence, these programs needed to build both a factory to stamp out a new product and a new distribution mechanism. A first generation infrastructure now exists, but it is not yet optimized to produce and deliver tests as efficiently as possible. Right now, there's no question about it: for these programs, assessment by computer costs far more than assessment by paper.

If we have learned anything from the history of innovation, it is that new technologies are often initially far too expensive for mass use. That was true of the automobile, telephone service, commercial aviation, and the personal computer, among many other innovations. For example, in 1930 the cost of a three-minute telephone call from New York to London was $250 (in 1990 dollars). By 1995, the cost had dropped to under $1 (World Bank, 1995, cited in Cairncross, 1997, p. 28). As a second instance, when the IBM Personal Computer was introduced in 1981, it cost around $5,000. At the time, the median family income in the United States was on the order of $25,000, so that a computer cost about 20% of the average family's earnings—not very affordable. At this writing, the cost of a computer with many times greater capability is a little more than $500 and the median income is closer to $55,000. (Note 9) A computer now costs about 1% of average income. (Note 10)

When a promising new technology appears, individuals and institutions invest, allowing the technology to evolve and a supporting infrastructure to develop. Over the course of that development, failures inevitably occur. Eventually, the technology either dies or becomes commercially viable—that is, efficient enough.

So, who's investing in CBT? At this point, it's an impressive list including non-profit testing agencies, for profit-testing companies, school districts, state education departments, government agencies, and companies with no history in testing at all. The list includes ACT, the Bloomington (MN) Public Schools, CITO (the Netherlands), the College Board, CTB/McGraw-Hill, Edison Schools, ETS, Excelsior College (formerly Regents College), Harcourt Educational Measurement, Heriot-Watt University (Scotland), Houghton-Mifflin, Microsoft, the National Board of Medical Examiners, the National Institute for Testing and Evaluation (Israel), NCS Pearson, the Northwest Evaluation Association, the Oregon Department of Education, the Qualifications and Curriculum Authority (Great Britain), Thomson Corporation, the University of Cambridge Local Examinations Syndicate (UCLES), the U.S. Armed Forces, Vantage Technologies, and the Victoria (Australia) Board of Studies. These organizations are producing tests for postsecondary admissions, college course placement, course credit, school accountability, instructional assessment, and professional certification and licensure (see the Appendix for details.) In concert, they already administer something on the order of 10 million computerized tests each year. (Note 11)

Why are these organizations investing? I think it's because they believe that technology-based assessment will eventually achieve important economies over paper and that, fundamentally, assessment will benefit. But I also think it's because they don't want to become *Britannica*. That is, they see improvements in the business and substance of assessment which, if they fail to embrace, will lead them to the same fate as that encyclopedia publisher.

CBT as a Disruptive Technology

But as the case of admissions testing suggests, the road to improvement may be a difficult one since CBT might not be a typical innovation. Christensen (1997) distinguishes between two types of innovation, called *sustaining* and *disruptive* technologies. Sustaining technologies enhance the performance of established products in ways that mainstream customers have traditionally valued. Historically, most technological advances in any given industry have been sustaining ones (e.g., in the personal computer industry, faster chips and bigger, higher-resolution monitors). Occasionally, disruptive technologies emerge. Companies introduce these technologies hoping their features will provide competitive edge. However, these features characteristically overshoot the market, giving customers more than they need or are willing to pay for. Thus, disruptive technologies result in *worse* product performance, at least in the near-term, on key dimensions in a company's established markets.

Interestingly, a few fringe customers typically find a disruptive technology's new features attractive. In these niche markets, such technology may thrive. If and when it advances to the level and nature of performance demanded in the mainstream market, the new technology can invade it, rapidly knocking out the traditional technology and its dependent practitioners. Remember *Britannica*.

CBT has many of the characteristics of a *disruptive* technology. Established testing organizations are applying it in their mainstream markets, most notably postsecondary admissions. This innovation was introduced, in good part, to provide competitive edge through features like the ability to take a test at one's convenience and to get score reports immediately. As it turned out, these features overshot the market. At least initially, registrations for continuously-offered computer-based admissions tests mirrored those for fixed-date administrations, suggesting that scheduling convenience was not a highly valued feature in the market of the time. Moreover, examinees were dissatisfied with losing some of the features of paper exams, including the ability to proceed through the test nonlinearly, the option to review the scoring of items actually taken, and the low cost (Perry, 2000).

Although it encountered difficulty in the mainstream admissions testing market, CBT found more rapid acceptance in the niches. One example is information technology (IT) certification, which individuals pursue to document their competence in some computer-related proficiency. In 1999, over three million examinations in 25 languages were administered in this market (Adelman, 2000). Most of these tests were delivered on computer and most were offered on a continuous basis. Three delivery vendors provided the bulk of examinations: CAT, Inc. (a subsidiary of Houghton-Mifflin), Prometric (a subsidiary of Thomson Corporation), and Vue (a subsidiary of NCS Pearson). Together, these vendors operated some 5,000 testing centers in 140 countries. As of June, 2000, over 1.9 million credentials had been awarded, most for Microsoft or Novell technologies.

Why is the CBT of today so well suited to this market niche? Let's start by asking what features a testing product must have to succeed in this niche. First, it must be continuously offered because these test candidates build technology skill on their own schedules—at home or on the job, very often through books or online learning. These individuals want to test when they are ready, not when the testing companies are. Second, such a test must generally be offered on computer since technology use is the essence of the certification.

What are the financial considerations associated with serving this market? One consideration is whether the test fee can cover the cost of assessment. As it turns out, this market is less price-sensitive than postsecondary admissions. Why? With IT testing, employers pay the fee for over half the candidates (Adelman, 2000). In addition, certified employees command a substantial salary premium (4-14%), which makes examinees more willing to absorb the higher fees that CBT currently requires. A second consideration is that security is not as critical as in admissions testing, so large item pools are not needed, reducing production cost. Lower security is tolerable because if an individual appears on the job with a dishonestly obtained credential but without the required skill, he or she will not last. Finally, test volume is self-replicating: there are many repeat test takers because information technology changes rapidly, so skills must be updated constantly. From an innovation perspective, then, IT certification may be one context in which the CBT of today can flourish and develop to better meet the needs of other assessment markets.

So why do industry leaders tend to fail with disruptive technology while fringe players succeed? Industry leaders often fail precisely because they attempt to introduce disruptive technologies into major markets before it's time (Christensen, 1997). Because niche markets are often too small to be of interest, leaders do not pursue those opportunities to refine the technology. Instead, they give up, having run out of resources or credibility. Making a disruptive technology work requires iteration and iteration means failure. Because they risk neither large resources nor reputations in the mainstream market, it is the fringe players who can fail early, often, and inexpensively enough to eventually challenge and overtake the industry leaders.

Toward the Technology Based Assessment of Tomorrow

Are there other niche markets in which CBT might evolve? One such niche may be online learning. If we believe the Web-Based Education Commission (Kerrey & Isakson, 2000), online learning will become a major enterprise, especially for the lifelong updating of skills. In this market, institutions will be less concerned with questions of who gets *in* and more with who gets *out*, and what it is they have to do to get out (Messick, 1999). Why? Because once hired, businesses are becoming more concerned with what employees know and can do, and less with where they went to school. Similarly, individuals are becoming more concerned with finding course offerings that meet their skill development goals and less with whether those offerings come from one institution or a half-dozen.

What's the assessment need? First, it is for *knowledge facilitation* and, second, for *knowledge certification;* that is, to help people develop their skills and then document that they've developed them. What's the assessment challenge? The challenge is to figure out how to design and deliver embedded assessment that provides instructional support and that globally summarizes learning accomplishment. In other words, the challenge is to combine richness with reach to achieve mass customization—use the Internet's ability to deliver the richness of customized assessment to reach a mass audience.

Can assessment be customized? In very rudimentary ways, it already is. Certainly, we can dynamically adapt along a global dimension, as is done in many of today's computerized tests. But as we move assessment closer to instruction, we should eventually be able to adapt to the interests of the learner and to the particular strengths and weaknesses evident at any particular juncture, as intelligent tutors now do (e.g., Schulze, Shelby, Treacy, & Wintersgill, 2000). Likewise, we should be able to customize feedback to describe the specific proficiencies the learner evidenced in an instructional sequence.

But perhaps the most far-reaching customization of assessment will come through modular online courses, whereby an instructor—or even a sophisticated learner—assembles a series of components into a unique offering. The Department of Defense (DOD) has taken a significant step through the Sharable Courseware Object Reference Model (SCORM) (www.adlnet.org). SCORM is to embody specifications and guidelines providing the foundation for how DOD will use technology to build and operate the learning environment of the future. SCORM will allow mixing and matching of learning segments to create lower cost, reusable training resources. (Note 12) If embedded assessment can be built into course modules following a similar set of conventional specifications, the assessment too will be customized by default.

Conclusion

Whether for postsecondary admissions, school and student accountability, or national policy, large-scale assessment must be reinvented. Reinvention is not an option. If we do not reinvent it, much of today's paper-based testing will become an anachronism—"yesterday's testing technology," in the words of the Web-Based Education Commission (Kerrey & Isakson, 2000)—because it will be inconsistent with what and how students learn.

This reinvention must occur along both business and substantive lines. As educators, we often behave as if business considerations are unimportant, even distasteful. However, the business and substance of assessment are intertwined. Even for non-profit educational institutions—state education departments, federal agencies, schools, research organizations—providing quality assessment for a low cost matters. Using new technology to do assessment faster and cheaper can free up the resources to do assessment better.

We will be able to do assessment better because advances in technology, cognitive science, and measurement are laying the groundwork to make reinvention a reality. Whereas the contributions of cognitive and measurement science are in many ways more fundamental than those of new technology, it is new technology that is pervading our society. My thesis, therefor, is that new technology will be the primary facilitating factor precisely because of its widespread societal acceptance. (Note 13) In the same way that the Internet is already helping to revolutionize commerce, education, and even social interaction, this technological advance will help revolutionize the business and substance of large-scale assessment. It will do so by allowing richness with reach—that is, mass customization on a global scale—as never before. However, as the history of innovation suggests, this reinvention won't come immediately, without significant investment, or without setback. With few exceptions, we are not yet ready for large-scale assessment via the Internet (at least in our schools). However, as suggested above, this story is not so much about today. It really is about tomorrow.

Notes

This article is based on a paper presented at the annual conference of the International Association for Educational Assessment (IAEA), Jerusalem, May 2000.

I appreciate the helpful comments of Isaac Bejar, Henry Braun and Drew Gitomer on an earlier draft of this manuscript.

1. The Internet takes advantage of many such standards, including Internet Protocol (IP) for transmitting packets of information; Transmission Control Procotol (TCP/IP) for verifying the contents of those packets; HyperText Transfer Protocol (HTTP) for transferring web-pages; and HyperText Markup Language (HTML) and Extensible Markup Language (XML) for representing structured documents and data on the Web. XML provides a significant advance over HTML in that it allows for the representation of unlimited classes of documents. Leadership in developing and implementing the many standards used by the Internet is provided by the World Wide Web Consortium (www.w3.org). For more on Internet standards, see their website or see Green (1996), who gives a more basic introduction.
2. According to Neilsen//NetRatings, 56% of U.S. households had Internet access as of November 2000 ("Internet access tops 56 percent," 2000).
3. And it works. eBay is reported to be the most successful company in cyberspace, with 22.5 million registered users and 2000 revenues of $430 million (Cohen, 2001). Why? It has none of the costs of retailing: No buying, no warehousing, no shipping, no returns, no overstock.
4. A recent, but potentially significant, addition to this population is the U.S. Army. In July, 2000, Secretary of the Army, Louis Caldera, announced a 600 million dollar program to allow any interested soldier to take college courses over the Internet at little or no cost (Carr, 2000b).
5. A second, perhaps more interesting, example is Florida's Daniel

Jenkins Academy, where students physically attend but take all academic courses *on-line* from *off-site* teachers (Thomas, 2000).

6. Russell has conducted several studies on the mismatch between learning and testing methods in writing (e.g., Russell & Plati, 2001). The repeated result is that the writing proficiencies of students who routinely use word processors are underestimated by paper-and-pencil tests.
7. The Teaching, Learning, and Computing—1998 survey provides similar data (Anderson & Ronnkvist, 1999). This survey, conducted using a national probability sample in Spring 1999, reports Internet access in 90% of schools and at least medium-speed, dedicated connections in 57%.
8. Developing a technology infrastructure and integrating into the e-commerce network may, in fact, help jump-start the growth required to deal with the serious problems of public health, education, and welfare that these countries typically face (Friedman, 2000).
9. The median income for a family of four in 1981 was $26,274 (U.S. Census Bureau, 2001). For 1998, it was $56,061.
10. Price and quality-adjusted data tell a similar story. In 1983, the quality-adjusted cost of a personal computer in constant 1996 dollars was $1098 (D. Wasshausen, personal communication, April 13, 2000). By 1996, the cost of a PC, holding quality constant, was $100, less than a tenth of the 1983 cost. By 1999, that quality-adjusted PC had further deflated to $29.
11. I based this estimate on unduplicated volumes claimed by Thomson Prometric (www.prometric.com), Vantage Technologies (www.intellimetric.com/index.html), and the U.S. Armed Forces (A. Nicewander, personal communication, November 2, 2000). These three organizations *alone* claim some 8.5 million tests annually. These tests include both high-stakes and low-stakes assessments.
12. SCORM is being built upon the work of the IMS Global Learning Consortium (IMS) (www.imsproject.org/aboutims.html). IMS is developing open specifications for facilitating distributed learning activities such as locating and using educational content, tracking learner progress, reporting learner performance, and exchanging student records between administrative systems. Both IMS and SCORM incorporate XML (see note 1 above).
13. That the largest *facilitating* factor will be technological is not to say that we should necessarily let technology drive the substance of assessment. We shouldn't.

References

ACT and EDS alliance to expand the nation's testing and training opportunities. (1999, June 8). *ACT Newsroom* [On-line]. Available: www.act.org/news/releases/1999/06-08-99.html

Adelman, C. (2000). A parallel postsecondary universe: The certification system in information technology. Washington, D.C.: Office of Educational Research and Improvement, U.S. Department of Education. Available: www.ed.gov/pubs/ParallelUniverse/

Anderson, R. E., & Ronnkvist, A. (1999). The presence of computers in American Schools. Irvine, CA: Center for Research on Information Technology and Organizations. Available: www.crito.uci.edu/tlc/findings/computers_in_american_schools/

Ball, S. (1999). Measurement and the culture of education: The story of VSAM. *Educational Measurement: Issues and Practice, 18(2),* 50-51.

Barton, P. E. (1999). What jobs require: Literacy, education, and training, 1940-2006. Princeton, NJ: Policy Information Center, Educational Testing Service. Available: www.ets.org/research/pic

Bennett, R. E. (1998). Reinventing assessment: Speculations on the future of large-scale educational testing. Princeton, NJ: Policy Information Center, Educational Testing Service. Available: www.ets.org/research/pic/bennett.html

Bennett, R. E. (1999). Using new technology to improve assessment. *Educational Measurement: Issues and Practice, 18(3),* 5-12.

Bennett, R. E., Goodman, M., Hessinger, J., Ligget, J., Marshall, G., Kahn, H., & Zack, J. (1999). Using multimedia in large-scale computer-based testing programs. *Computers in Human Behavior, 15,* 283-294.

Big fish in a big pool. (1999). *TIME Digital*, December 2.

Burstein, J., Braden-Harder, L., Chodorow, M., Hua, S., Kaplan, B., Kukich, K., Lu, C., Nolan, J., Rock, D., & Wolff, S. (1998). Computer analysis of essay content for automated score prediction (RR-98-15). Princeton, NJ: Educational Testing Service.

Cairncross, F. (1997). *The death of distance: How the communications revolution will change our lives*. Boston, MA: Harvard Business School Press.

Carr, S. (1999, December 10). 2 more universities start diploma-granting virtual high schools. *The Chronicle of Higher Education,* p. A49.

Carr, S. (2000a, March 24). Cornell creates a for-profit subsidiary to market distance education programs. *The Chronicle of Higher Education,* p. A47.

Carr, S. (2000b, August 18). Army bombshell rocks distance education. *The Chronicle of Higher Education,* p. A35.

Carr, S., & Young, J. R. (1999, October 22). As distance learning boom spreads, colleges help set up virtual high schools. *The Chronicle of Higher Education,* p. A55.

Cerf, V. (1993). How the Internet came to be. In B. Aboba (Ed.), *The online user's encyclopedia*. New York: Addison-Wesley. Available: http://www.bell-labs.com/user/zhwang/vcerf.html

Christensen, C. M. (1997). *The innovator's dilemma: When new technologies cause great firms to fail*. Boston, MA: Harvard University Press.

Church, G. J. (1999). The economy of the future? *TIME*, 154(14). Available: http://www.time.com/time/magazine/article/0,9171,31522,00.html

Cisco Systems, Inc. (2000). *Discover all that's possible on the Internet: 2000 annual report*. San Jose, CA: Cisco Systems, Inc. Available: www.cisco.com/warp/public/749/ar2000

Clauser, B. E., Margolis, M. J., Clyman, S. G., & Ross, L. P. (1997). Development of automated scoring algorithms for complex performance assessments: A comparison of two approaches. *Journal of Educational Measurement, 34,* 141-161.

Cohen, A. (1999). The attic of e. *TIME*, 154(26). Available: http://www.time.com/time/magazine/article/0,9171,36306-1,00.html

Cohen, A. (2001). eBay's bid to conquer all. *TIME, 157(5)*, 48-51.

Department of Education. (2000). Request for proposals for the technology enhanced student assessment system. Salem, OR: Department of Education. Available: www.ode.state.or.us/asmt/develop/rfptesa.htm

Dunn, S. L. (2000). The virtualizing of education. *The Futurist, 34(2),* p.34-38

Early test prep. (1999). ABC News.com [On-line]. Available: http://abcnews.go.com/sections/tech/DailyNews/testing991020.html

Education prognosis 1999. (1999, January 11). *Business Week*, 132-133.

Evans, P., & Wurster, T. S. (2000). *Blown to bits: How the economics* *of information transforms strategy*. Boston, MA: Harvard Business School Press.

FCC: E-rate subsidy funded at $2.25 billion cap. (2000). *What Works in Teaching and Learning, 32(8)*, p. 8.

Fernandez, S. M. (2000). Latin America logs on. *TIME, 155(19)*, B2-B4.

Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist, 39*, 193-202.

Friedman, T. L. (2000). *The Lexus and the olive tree: Understanding globalization*. New York: Anchor Books.

Gershenfeld, N. (1999). *When things start to think*. New York: Holt.

Gibney, Jr., F. (2000). Enron plays the pipes. *TIME, 156(9)*, 38-39.

Gitomer, D. H., Mislevy, R. J., & Steinberg, L. S. (1995). Diagnostic assessment of troubleshooting skill in an intelligent tutoring system. In P. D. Nichols, S. F. Chipman, R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 72-101). Hillsdale, NJ: Erlbaum.

Glaser, R. (1991). Expertise and assessment. In M. C. Wittrock & E. L. Baker (Eds.), *Testing and cognition* (pp. 17-30). Englewood Cliffs, NJ: Prentice-Hall.

Global Reach. (2000). Global Internet statistics (by language). Available: www.glreach.com/globstats/index.php3

Green, C. (1996). An introduction to Internet protocols for newbies. Available: www.halcyon.com/cliffg/uwteach/shared_info/internet_protocols.html

Grice, C. (2000). Wireless handhelds will rule the day, PC execs predict. *CNET News.com* [On-line]. Available: http://news.cnet.com/news/0-1004-200-1560446.html

Henry, D., Buckley, P., Gill, G., Cooke, S., Dumagan, J., Pastore, D., & LaPorte, S. (1999). The emerging digital economy II. Washington, D.C.: U.S. Department of Commerce. Available: www.ecommerce.gov/ede/ede2.pdf

How many online? (2000, December 20). Nua Internet Surveys. Available: www.nua.ie/surveys/how_many_online/index.html

Internet access tops 56 percent in U. S., according to Neilsen//NetRatings. (2000, December 18). Available:

http://209.249.142.22/press_releases/PDF/pr_001215.pdf

Internet domain survey host count. (2000). Internet Software Consortium. Available: www.isc.org/ds/hosts.html

Kennard, W. E. (2000, January). E-rate: A success story. Presentation at the Educational Technology Leadership Conference- -2000, Washington, D.C.

Kerrey, B., & Isakson, J. (2000). The power of the Internet for learning: Moving from promise to practice. (Report of the Web-based Education Commission). Washington, D.C.: Web-Based Education Commission. Available: http://interact.hpcnet.org/webcommission/index.htm

Landler, M. (1995, May 16). Slow-to-adapt *Encyclopaedia Britannica* is for sale. *New York Times*, D1, D22.

Levine, A. (2000a, March). The remaking of the American university. Paper presented at the Blackboard Summit, Washington, D. C.

Levine, A. (2000b, March 13). The soul of a new university. *New York Times*, p. 21.

Lewis, L., Snow, K., Farris, E., Levin, D., & Greene, B. (1999). Distance education at postsecondary education institutions: 1997-1998 (NCES Statistical Analysis Report 2000-013). Washington, D.C.: National Center for Education Statistics. Available: http://nces.ed.gov/pubs2000/2000013.pdf

Maney, K. (2000). E-novel approach promises new chapter for book lovers. *USA Today, 18(169),* 8A-9A.

Melcher, R. A. (1997). Dusting off the *Britannica*: A new order has digital dreams for the august encyclopedia. *Business Week Online*. Available: www.businessweek.com/1997/42/b3549124.htm

Mendels, P. (1999). The leading issues of '99? Wired schools and accreditation. *The New York Times On the Web* [On-line]. Available: www.nytimes.com/library/tech/99/12/cyber/education/29education.html

Messick, S. (1999). Technology and the future of higher education assessment. In S. Messick (Ed.), *Assessment in higher education: Issues of access, student development, and public policy* (pp. 245-254). Hillsdale, NJ: Erlbaum.

Moe, M. T., & Blodget, H. (2000). *The knowledge web: People power—Fuel for the new economy*. San Francisco: Merrill Lynch.

National Center for Education Statistics. (2000). Stats in brief: Internet access in US public schools and classrooms: 1994-99. Washington, DC: US Department of Education, Office of Research and Improvement.

NCS secures rights to iPaper electronic technology in testing and education market. (2000, July 11). Minneapolis, MN: National Computer Systems (NCS). Available: www.ncs.com/ncscorp/top/news/000711.htm

Negroponte, N. (1995). *Being digital*. New York: Vintage.

Odendahl, N. (1999, April). Online delivery and scoring of constructed-response assessments. Paper presented at the annual meeting of the American Educational Research Association, Montreal.

Olsen, F. (2000, February 18). A UCLA professor and net pioneer paves the way for the next big thing. *The Chronicle of Higher Education, 46*.

Pellegrino, J. W., Jones, L. R., & Mitchell, K. J. (1999). *Grading the nation's report card*. Washington, D.C.: National Academy Press.

Perry, J. (2000). Digital tests spark controversy: Critics say revamped exams limit the options to challenge a score. *Online US News* [On-line]. Available: www.usnews.com/usnews/edu/beyond/grad/gbgre.htm

Poised to go global: Accuplacer online sales soar. (2000, April). *The Bulletin Board, 5(9)*, 5.

Report: College Net use growing. (2000, March 16). USA Today.com [On-line]. Available: www.usatoday.com/life/cyber/tech/cth566.htm

Riley, R. W., Holleman, F. S., & Roberts, L. G. (2000). e-Learning: Putting a world-class education at the fingertips of all children (The national educational technology plan). Washington, D.C.: U.S. Department of Education. Available: www.ed.gov/Technology/elearning/e-learning.pdf

Russell, M., & Plati, T. (2001). Effects of computer versus paper administration of a state-mandated writing assessment. *Teachers College Record*. Available: www.tcrecord.org/Content.asp?ContentID=10709 .

Schmidt, P. (2000, January 21). Judge sees no bias in Texas test for high-school graduation. *Chronicle of Higher Education*, p. A27.

Schulze, K. G., Shelby, R. N., Treacy, D. J., & Wintersgill, M. C. (2000, April). Andes: A coached learning environment for classical Newtonian physics. In *Proceedings of the 11th International*

Conference on College Teaching and Learning, Jacksonville, FL. Available: www.pitt.edu/~vanlehn/icctl.pdf

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29(7)*, 4-14.

Singley, M. K., & Bennett, R. E. (in press). Item generation and beyond: Applications of schema theory to mathematics assessment. In S. Irvine & P. Kyllonen (Eds.), *Item generation for test development*. Hillsdale, NJ: Erlbaum.

State of Georgia. (2001). Request for proposal number 41400-026-0000000031. Available: http://www2.state.ga.us/Departments/doas/procure/rfp/rfp-41400-026-0000000031.doc

Thomas, K. (2000, April 6). One school's quantum leap. *USA Today*. 1A. Available: www.usatoday.com/usatonline/20000406/2117463s.htm

Tulloch, J. B. (2000). Sophisticated technology offers higher education options. *T.H.E. Journal* [On-line]. Available: www.thejournal.com/magazine/vault/A3165.cfm .

U.S. Census Bureau. (2001). Median income for 4-person families, by state. Available: www.census.gov/ftp/pub/hhes/income/4person.html

U.S. Department of Commerce. (2000). Falling through the Net: Toward digital inclusion. Available: www.esa.doc.gov/fttn00.pdf

U.S. dominance seen slipping in Internet use, commerce. (2001). Cyberatlas: The Big Picture Geographics. Available: http://cyberatlas.internet.com/big_picture/geographics/article/0,,5911_377801,00.html

Virginia Department of Education. (Undated). Demonstrating success: A statewide web-based Standards of Learning technology and on-line testing initiative (Request for proposal # RFP-WEB2000). Richmond, VA: Virginia Department of Education. Available: www.pen.k12.va.us/VDOE/Technology/soltech/rfp/rfpweb2000.pdf

Whalen, S. J., & Bejar, I. I. (1998). Relational databases in assessment: An application to online scoring. *Journal of Educational Computing Research, 18,* 1-13.

Appendix: Some Organizations Investing in Computer-Based Testing

ACT, Inc. In partnership with EDS, ACT, Inc. is establishing a nationwide network of electronic testing and training centers. These centers will provide computer-delivered certification and licensure tests for the trades and professions; a computerized measure of workplace skills to guide training decisions; and computerized educational and career guidance. More than 250 ACT Centers are expected to be operational by the end of 2001 ("ACT and EDS," 1999). ACT also offers a computerized placement test for post-secondary institutions to use in determining whether entering students need assignment to remedial or developmental courses in mathematics, reading, writing, and English-as-a-second-language (www.act.org/compass/).

Bloomington (MN) Public Schools. This district was reportedly the first in the US to do its math and reading testing exclusively via computer ("Early test prep," 1999). Bloomington uses an intranet-delivered computer-adaptive test designed by the Northwest Evaluation Association (see entry below) (www.bloomington.k12.mn.us/Staff_Resources/ Office_of_Research_and_Evaluat/CALT_Technical_Description /calt_technical_description.htm).

CITO. CITO, the measurement organization of the Netherlands, has developed a computerized adaptive test, WisCat, for placement in adult education. WisCat is used by approximately half the vocational training institutes in the Netherlands (Verschoor, personal communication, November 7, 2000).

College Board. The College Board offers Accuplacer, an adaptive placement test that can be delivered over the Internet for use in postsecondary institutions (www.collegeboard.org/accuplacer/html/ accupla1.html). Last year, over 2 million exams were administered ("Poised to go global," 2000), probably making Accuplacer the largest volume CBT in the world. By July 2001, the Board will also be offering its entire College Level Examination Program (CLEP) on computer: over 30 tests designed to allow individuals to get college credit for knowledge gained outside of school (www.collegeboard.com/clep/clepcntr/html/tc0 01.html).

CTB/McGraw-Hill. This company offers a PC version of the Test of Adult Basic Education, a measure of reading, mathematics, language, and spelling skills used in adult literacy programs (www.ctb.com/products_services/tabe/ index.html).

Edison Schools. This for-profit company manages 113 public schools with a total enrollment of 57,000 students. Edison recently introduced its Benchmark Assessment System, designed to provide teachers with ongoing, instructionally relevant information about the progress of their 2nd to 8th grade students. These computerized assessments in reading, math, writing, and language arts will be administered over 1 million times during the 2000-2001 academic year (www.intellimetric.com/when.newstoday0.html).

Educational Testing Service (ETS). In the 1999-2000 year, ETS administered over a million tests on computer for the GRE, GMAT, and TOEFL programs. In addition, a variety of licensure and certification examinations were given through ETS' Chauncey Group International subsidiary (www.ets.org/cbt/index.html). A second subsidiary, ETS Technologies, markets automated scoring services for computer-delivered writing tests (www.etstechnologies.com).

Excelsior College (formerly Regents College). Excelsior computerized exams allow adults to demonstrate their college-level knowledge in the arts and sciences, business, education, and nursing. Students may use these exams for advanced placement and exemption from course requirements, or to obtain Excelsior College degrees (www.excelsiorcollege.com).

Harcourt Educational Measurement (HEM). HEM offers a web-based version of the Stanford Writing Assessment Program in English and 15 foreign languages for use in grades 3 through 12 (www.hbem.com/trophy/achvtest/index.htm).

Heriot-Watt University. This Edinburgh (Scotland) institution uses web-based testing extensively in its on-campus and distance learning courses for both self-assessment and final examinations (http://flex-learn.ma.hw.ac.uk/ info.html). The success of the technology and its spread to other Scottish universities led to a spin off, Web4Test.Ltd, to commercialize the technology (http://web4test.com/comp.html).

Houghton-Mifflin. CAT, Inc., a subsidiary, offers computer-based tests for credentialing, training, and employment (http://catinc.com).

Microsoft. Microsoft develops computer-based tests to certify individuals in many of its software products (www.microsoft.com/trainingandservices/ default.asp?PageID=mcp).

National Board of Medical Examiners (NBME). NBME develops the United States Medical Licensing Examination. *All* individuals wanting to be licensed to practice medicine in the U.S. must take this computer-based test, including a section having clinical case simulations (www.usmle.org/home.htm).

National Institute for Testing and Evaluation (NITE). This Israeli measurement organization offers a college placement test similar to those marketed by the College Board and ACT, Inc.

NCS Pearson (formerly National Computer Systems). Through its VUE subsidiary, NCS Pearson delivers tests for information technology certification, including those developed by Microsoft, as well as for Cisco Systems, Novell, and IBM (www.vue.com).

Northwest Evaluation Association (NWEA). NWEA has its Measures of Academic Progress, which assesses growth in reading, mathematics, language, and science. The web-delivered version of this test is used in 1,100 schools in 90 school districts (M. Patterson, personal communication, October 23, 2000) (www.nwea.org/PRODUCTS/MAP.htm).

Oregon Department of Education, Virginia Department of Education, and Georgia Department of Education. These state departments are each developing systems for web-based assessment designed to serve both instructional and accountability purposes (www.ode.state.or.us/asmt/develop/rfptesa.htm , www.pen.k12.va.us/VDOE/Technology/soltech/ rfp/rfpweb2000.pdf, http://www2.state.ga.us/Departments/doas/ procure/rfp/rfp-41400-026-0000000031.doc). Virginia plans to begin delivering its computer assessments to all state high schools by 2003.

Qualifications and Curriculum Authority (QCA). This organization, responsible for British national assessment, is developing the World Class Tests. These exams are intended to recognize the achievements of gifted and talented children worldwide in mathematics and problem solving. The tests, which will be largely computer-delivered, debut operationally in November 2001 (www.qca.org.uk/ca/tests/wct/about_the_tests.asp).

Question Mark Corporation. Question Mark sells software for authoring and delivering web-based tests (www.questionmark.com/home.htm).

Thomson Corporation. In 1999, Thomson's Prometric subsidiary delivered over four *million* tests for 140 organizations, including ETS, Excelsior College, Microsoft, and the National Board of

Medical Examiners (www.prometric.com). Thomson also recently announced its intention to purchase Harcourt's Assessment Systems, Inc., which administers computerized tests for occupational and professional licensure and certification, as well as for employment (www.asisvcs.com).

University of Cambridge Local Examinations Syndicate (UCLES). UCLES offers a computerized-adaptive version of its Business Language Testing Service (BULATS) on CD-ROM. BULATS helps organizations assess the language skills of job applicants, trainees, and employees. The test is available in English, French, German, and Spanish (www.bulats.org/suite.cfm). UCLES is developing several other computerized language tests, including a version of its International English Language Testing System (IELTS).

U.S. Armed Forces. Since the early 1990s, the U.S. Armed Forces has been administering *its* admissions test, the Armed Services Vocational Aptitude Battery, on computer. This adaptive test is given about 450,000 times per year. Because the test is shorter than its paper-and-pencil counterpart, processing can be completed in one day, saving the armed services considerable cost in housing applicants (A. Nicewander, personal communication, November 2, 2000).

Vantage Technologies. This small, Yardley (PA) company claims to be the largest provider of computer-based tests (www.intellimetric.com/index.html). Depending upon what one includes, that claim may be correct. Among other things, Vantage administers Accuplacer for the College Board and the Benchmark Assessment System for Edison Schools. In addition, it will be delivering state assessments via the web for the Oregon Department of Education.

Victoria, Australia Board of Studies. Victoria is beginning to deliver state-wide achievement tests via the Internet (Ball, 1999).

About the Author

Randy Elliot Bennett
Educational Testing Service
Princeton, NJ 08541
Email: rbennett@ets.org

Randy Bennett is Distinguished Presidential Appointee at Educational Testing Service in Princeton, NJ, a nonprofit organization dedicated to research and service in educational measurement. Dr. Bennett began his employment at ETS in 1979. Since the 1980's, he has conducted research on the applications of technology to testing and teaching, on new forms of assessment, and on the assessment of students with disabilities. Dr. Bennett's work on the use of new technology to improve assessment has included research on presenting and scoring open-ended test items via computer, on multimedia in testing, and on generating test items automatically. Dr. Bennett is the editor or author of seven books and many other publications including a widely-cited monograph, "Reinventing Assessment: Speculations on the Future of Large-Scale Educational Testing" (http://www.ets.org/research/pic/bennett.html). He has made presentations on this and related topics throughout the world. Dr. Bennett is currently leading a series of studies designed to lay the groundwork for introducing computerized testing to the U.S. National Assessment of Educational Progress.

Copyright 2001 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Sherman Dorn
University of South Florida

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
hmwkhelp@scott.net

Thomas F. Green
Syracuse University

Alison I. Griffith
York University

Arlen Gullickson
Western Michigan University

Ernest R. House
University of Colorado

Aimee Howley
Ohio University

Craig B. Howley
Appalachia Educational Laboratory

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Benjamin Levin
University of Manitoba

Thomas Mauhs-Pugh
Green Mountain College

Dewayne Matthews
Western Interstate Commission for Higher Education

William McInerney
Purdue University

Mary McKeown-Moak
MGT of America (Austin, TX)

Les McLean
University of Toronto

Susan Bobbitt Nolen
University of Washington

Anne L. Pemberton
apembert@pen.k12.va.us

Hugh G. Petrie
SUNY Buffalo

Richard C. Richardson
New York University

Anthony G. Rud Jr.
Purdue University

Dennis Sayers
Ann Leavenworth Center for Accelerated Learning

Jay D. Scribner
University of Texas at Austin

Michael Scriven
scriven@aol.com

Robert E. Stake
University of Illinois—UC

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

other vols. | abstracts | editors | board | submit | comment | subscribe | search

This article has been retrieved 1658 times since February 22, 2001

other vols. | abstracts | editors | board | submit | comment | subscribe | search

Education Policy Analysis Archives

Volume 9 Number 6 **February 22, 2001** **ISSN 1068-2341**

A peer-reviewed scholarly journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2001, the **EDUCATION POLICY ANALYSIS ARCHIVES**.
Permission is hereby granted to copy any article
if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in
the *Current Index to Journals in Education*
by the ERIC Clearinghouse on Assessment
and Evaluation and are permanently archived
in *Resources in Education*.

Teacher Test Accountability: From Alabama to Massachusetts

Larry H. Ludlow
Boston College

Abstract

Given the high stakes of teacher testing, there is no doubt that every teacher test should meet the industry guidelines set forth in the *Standards for Educational and Psychological Testing*. Unfortunately, however, there is no public or private business or governmental agency that serves to certify or in any other formal way declare that any teacher test does, in fact, meet the psychometric recommendations stipulated in the *Standards*. Consequently, there are no legislated penalties for faulty products (tests) nor are there opportunities for test takers simply to raise questions about a test and to have their questions taken seriously by an impartial panel. The

> purpose of this article is to highlight some of the psychometric results reported by National Evaluation Systems (NES) in their *1999 Massachusetts Educator Certification Test (MECT) Technical Report*, and more specifically, to identify those technical characteristics of the MECT that are inconsistent with the *Standards*. A second purpose of this article is to call for the establishment of a standing test auditing organization with investigation and sanctioning power. The significance of the present analysis is twofold: a) psychometric results for the MECT are similar in nature to psychometric results presented as evidence of test development flaws in an Alabama class-action lawsuit dealing with teacher certification (an NES-designed testing system); and b) there was no impartial enforcement agency to whom complaints about the Alabama tests could be brought, other than the court, nor is there any such agency to whom complaints about the Massachusetts tests can be brought. I begin by reviewing NES's role in *Allen v. Alabama State Board of Education*, 81-697-N. Next I explain the purpose and interpretation of standard item analysis procedures and statistics. Finally, I present results taken directly from the *1999 MECT Technical Report* and compare them to procedures, results, and consequences of procedures followed by NES in Alabama.

Teacher Test Accountability: From Alabama to Massachusetts

From its inception and continuing through present administrations, the Massachusetts Educator Certification Test (MECT) has attracted considerable public attention both regional and around the world (Cochran-Smith & Dudley- Marling, in press). This attention is due in part to two disturbing facts: 1) educators seeking certification in Massachusetts have generally performed poorly on the test, and 2) in many instances politicians have used these test results to assert, among other things, that candidates who failed are "idiots" (Pressley, 1998).

The purpose of the MECT is "to ensure that each certified educator has the knowledge and some of the skills essential to teach in Massachusetts public schools" (National Evaluation Systems, 1999, p. 22). The Massachusetts Board of Education has raised the stakes on the MECT by enacting plans to sanction institutions of higher education (IHEs) with less than an 80% pass rate for their teacher candidates (Massachusetts Department of Education, 2000). One consequence of this proposal is that most IHEs are considering requirements that the MECT be passed before students are admitted to their teacher education programs. In addition, Title II (Section 207) of the Higher Education Act of 1998 requires the compilation of state "report cards" for teacher education programs, which must include performance on certification examinations (U.S. Department of Education, 2000).

What all of this means is that poor performance on the MECT could prevent federal funding for professional development programs, limit federal financial aid to students, allow some IHEs be labeled publicly "low performing", and prove damaging at the state-level when states are inevitably compared to one another upon release of the Title II report cards in October 2001. Given the personal, institutional, and national ramifications of the test results, there is no question that the MECT should be expected to meet the industry benchmarks for good test development practice as set forth in the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999). At this time, however, there is no public or private business or governmental agency either within the Commonwealth of Massachusetts or nationally that can certify or in any other formal way declare that the MECT does (or does not), in fact, meet the psychometric recommendations stipulated in the *Standards.* The National Board on Educational Testing and Public Policy (NBETPP) serves as an "independent organization that monitors testing in the US" but even it does not function as a regulatory agency (NBETPP, 2000).

In addition to the absence of a national regulatory agency, many state departments of education do not have the professionally trained staff to answer directly technical psychometric questions. Nor do they usually have the expertise on staff to confront a testing company, which they have contracted, and demand a sufficient response to a technical question raised by outside psychometricians. Furthermore, even when a database with the candidates' item- level responses is available for internal analysis, a state department of education does not typically conduct rigorous disconfirming analyses, e.g. evidence of adverse impact. Thus, most state departments are largely dependent on whatever information testing companies decide to release. The public is then left with an inadequate accountability process.

One purpose of this article is to highlight some of the psychometric results reported by National Evaluation Systems in their *1999 MECT Technical Report* (NES, 1999). Specifically, this article identifies technical characteristics of the MECT that are inconsistent with the *Standards*. A second purpose of this article is to voice one more call for the establishment of a standing test auditing organization with powers to investigate and sanction (National Commission on Testing and Public Policy, 1990; Haney, Madaus & Lyons, 1993).

The significance of the present analysis is twofold. First, psychometric results reported by NES for the MECT are similar in nature to psychometric results entered as evidence of test development flaws in an Alabama class- action lawsuit dealing with teacher certification (*Allen v. Alabama State Board of Education*, 81-697-N). That suit was brought by several African-American teachers who charged, among other things, that "the State of Alabama's teacher certification tests impermissibly discriminate[d] against black persons seeking teacher certification;" the tests "[were] culturally biased;" and the tests "[had] no relationship to job performance" (*Allen*, 1985, p. 1048). Second, there was no impartial enforcement agency to whom complaints about the Alabama tests could be brought, other than the court, nor is there any such agency to whom complaints about the Massachusetts tests can be brought. These two points are linked in an interesting and troubling way--NES, the Massachusetts Educator Certification Tests contractor, was also the contractor for the Alabama Initial Teacher Certification Testing Program (AITCTP).

Some of the criticism of debates about teacher testing, teacher standards, teacher quality, and accountability suggests that arguments are, in part, ideologically, rather than empirically based (Cochran-Smith, in press). This may or may not be the case. This article, however, takes the stance that regardless of one's political ideology or philosophy about testing, the MECT is technically flawed. Furthermore, because of the lack of an enforceable accountability process, the public is powerless in its efforts to question the quality or challenge the use of this state-administered set of teacher certification examinations. In this article I argue that the consequences of high-stakes teacher certification examinations are too great to leave questions about technical quality solely in the hands of state agency personnel, who are often ill- prepared and under-resourced, or in the hands of test contractors, who may face obvious conflicts-of-interest in any aggressive analyses of their own tests.

In the sections that follow, I begin by reviewing NES's role in *Allen v Alabama*. Then I explain the purpose and interpretation of standard item analysis procedures and statistics. Finally I compare results taken directly from the *1999 MECT Technical Report* with statistical results entered as evidence of test development flaws in *Allen v Alabama*.

NES and the AITCTP

Allen, et al. v. Alabama State Board of Education, et al.

In January 1980, National Evaluation Systems was awarded a contract on a non-competitive basis for the development of the Alabama Initial Teacher Certification testing Program (AITCTP). Item writing for these tests began in the Spring of 1981, and the first administration of the tests took place on June 6, 1981. *Allen v Alabama* was brought just six months later on December 15th, 1981. The Allen complaint challenged the Alabama State Board of Education's requirement that applicants for state teacher certification pass certain standardized tests administered under the AITCTP. On October 14, 1983, class certification (Note 1) was granted, and the first trial was set for April 22, 1985. Subsequent to a pre-trial hearing on December 19, 1984 and "after substantial discovery was done,"(Note 2) an out-of-court settlement was reached on April 4, 1985. A Consent Decree was presented to the U.S. District Court April 8, 1985(Note 3). The Attorney General for the State of Alabama immediately "publicly attacked the settlement" (*Allen*, 1985, p. 1050), claiming that it was illegal. Nonetheless, the consent decree was accepted by the court October 25, 1985 (*Allen*, Oct. 25. 1985). A succession of challenges and appeals on the legality and enforceable status of the settlement resulted (Note 4). For example, on February 5, 1986, the district court vacated its October 25th order approving the consent decree (*Allen*, February 5, 1985, p. 76). While the plaintiffs appeal of the February 5th decision was pending at the 11th Circuit Court of Appeals, trial began in district court on May 5, 1986.

The AITCTP consisted of an English language proficiency examination, a basic professional studies examination, and 45 content-area examinations. The purpose of the examinations was to measure "specific competencies which are considered necessary to successfully teach in the Alabama schools" (*Allen*, Defendants' Pre-Trial Memorandum, 1986, p. 21). A pool of 120 items for each exam was generated--100 of which were scorable and mostly remained unchanged across the first eight administrations. Extensive revisions were incorporated into most of the tests at the ninth administration. By the start of the May 1986 trial the tests had been administered 15 times in all.

A team of technical experts (Note 5) for the plaintiffs was hired in November 1983 (prior to the ninth administration of the exams) to examine test development, administration, and implementation procedures. The team was initially unsure about the form of the sophisticated statistical analyses they assumed would have to be conducted to test for the presence of "bias" and "discrimination", the bases of the case. That is, the methodology for investigating what was then called "bias" and is now called "differential item functioning" was far from well established at that time (Baldus & Cole, 1980). Nevertheless, when the plaintiffs' team received the student-level item response data from the defendants, their first steps were to perform an "item analysis." Such an analysis produces various item statistics and test reliability estimates. These initial analyses produced negative point-biserial correlations. Although point-biserial correlations are explained in detail below, suffice it to say at this point that it was a surprise to find negative point-biserial correlations between the responses that examinees provided on individual items and their total test scores. Such correlations are not an intended outcome from a well-designed testing program.

These statistical results prompted a detailed inspection of the content, format, and answers for all the individual items on the AITCTP tests. Content analyses yielded discrepancies in the keyed correct responses in the NES test documents and the keyed correct responses in the NES- supplied machine scorable answer keys (i.e., miskeyed items were on the answer keys). This finding led to an inspection of the original NES in-house analyses which revealed that negative point-biserials for scorable items existed in their own records from the beginning of the testing program and continuing throughout the eighth administration without correction.

What this meant for the plaintiffs was that NES had item analysis results in their own possession which indicated that there were miskeyed items. Nonetheless they implemented no significant changes in the exams until they were faced with a lawsuit and plaintiffs' hiring of the testing experts to do their own analyses. The defendants argued that it was normal for some problems to go undetected or uncorrected in a large-scale testing program because the overall effect is trivial for the final outcome. The problem with that argument was that many candidates were denied credit for test items on which they should have received credit, and some of those candidates failed the exam by only one point. In fact, as the plaintiffs argued, as many as 355 candidates over eight administrations of the basic professional skills exam alone should have passed but were denied that opportunity simply because of faulty items that remained on the tests (Milman, 1986, p. 285). It should be noted here that these were items that even one of the state's expert witnesses for the defense admitted were faulty (Millman, 1986, p. 280).

Establishing that there were flawed items with negative point-biserial correlations was critical to the plaintiffs' case. The plaintiffs presented as evidence page after page of so-called "failure tables" (Note 6) with the names of candidates for each test whose answers were mis-scored on these faulty items. Based upon these failure tables, any argument from defendants that the mis-keyed items did not change the career expectations for some candidates would most likely have failed.

In the face of this evidence, the defendants argued at trial that

> ...the real disagreement is between two different testing philosophies. One of these philosophies would require virtual perfection under its proponents' rigid definition of that word. The other looks at testing as a constantly-developing art in which professional judgment ultimately determines what is appropriate in a particular case" (*Allen*, Defendant's Pre-trial Memorandum, 1986, p. 121-2).

Plaintiffs counter-argued

> "This case…is not a philosophical case at all. This case is a case on professional competence….this was an incompetent job, unprofessional, and as I said before, sloppy and shoddy, and in the case of the miskeyed items, unethical." (Madaus, 1986, p. 185).

Judge Thompson, in the subsequent *Richardson* decision which also involved the AITCTP, specifically agreed with plaintiffs on this point (*Richardson*, 1989, p. 821, 823, 825). Excellent reviews of the diametrically opposed plaintiff and defendant positions may be found in Walden & Deaton (1988) and Madaus (1990).

At the same time that this case was proceeding, the plaintiffs' appeal to reverse the vacating of the original settlement was granted prior to a decision in this trial (*Allen*, Feb. 5, 1986, p. 75). The U.S. Court of Appeals decided the district court should have enforced the consent decree (*Allen*, April 22, 1987)—which the district court so ordered on May 14, 1987 (*Allen*, May 14, 1987). Although the decision to uphold the original settlement was a positive ruling for the plaintiffs, it also was somewhat counter-productive for them because it was unexpectedly beneficial to NES at this stage in the proceedings. That is because the evidence presented above in *Allen v Alabama* was critical of the state and NES (NES was explicitly referred to in the court documents). Thus, NES's best hope for avoiding a written opinion critical of their test development procedures was if plaintiffs' appeal were to be upheld and the original settlement enforced, as it was. Then there would be no evidentiary record, no court ruling, and no legal opinion that would reflect badly upon the NES procedures. *Richardson v Lamar County Board of Education* (87-T-568-N) commenced, however, and the actions of NES and the Alabama State Board of Education were openly discussed and critiqued in the court's opinion of November 30, 1989 (though NES was not mentioned by name in the *Richardson*, 1989 decision).

Richardson v Lamar County Board of Education, et al.

Like *Allen v Alabama, Richardson v Lamar County* also addressed issues of the "racially disparate impact" of the AITCTP (*Richardson*, 1989, p. 808). The Honorable Myron H. Thompson again presided, and testimony from *Allen v Alabama* was admitted as evidence (*Richardson*, 1989). Although the defendants denied in the *Allen v Alabama* consent decree that the AITCTP tests were psychometrically invalid, and even though no decision was reached in the abbreviated *Allen v Alabama* trial, the State Board of Education did not attempt to defend the validity of the tests in *Richardson v Lamar* and, "in fact, it conceded at trial that plaintiff need not relitigate the issue of test validity" (*Richardson v Alabama State Board of Education*, 1991, p. 1240, 1246).

Judge Thompson's position on the test development process of NES was clearly stated: "In order to fully appreciate the invalidity of the two challenged examinations, one must understand just how bankrupt the overall methodology used by the State Board and the test developer was" (*Richardson*, 1989, p. 825, n. 37). While sensitive to the fact that "close scrutiny of any testing program of this magnitude will inevitably reveal numerous errors," the court concluded that these errors were not "of equal footing" and "the error rate per examination was simply too high" (*Richardson*, 1989, pp. 822- 24) Thus, none of the examinations that comprised the certification test possessed content validity because of five major errors by the test developer and the test developer had made six major errors in establishing cut scores (*Richardson*, 1989, pp. 821-25).

Case Outcomes in Alabama

The *Allen v Alabama* consent decree required Alabama to pay $500,000 in liquidated damages and issue permanent teaching certificates to a large portion of the plaintiff class (*Allen*, Consent Decree, Oct. 25, 1985, pp. 9-11). The decree also provided for a new teacher certification process. However, no new test was developed or implemented and the Alabama State Board of Education suspended the teacher certification testing program on July 12, 1988. In 1995 the Alabama State Legislature enacted a law requiring that teacher candidates pass an examination as a condition for graduation. Subsequently, another trial was held February 23, 1996 to decide the state's motions to modify or vacate the 1985 consent decree (*Allen*, 1997, p. 1414). Those motions were denied on September 8, 1997 (*Allen*, Sept. 8, 1997). Given the rigorous test development and monitoring conditions of the Amended Consent Decree, it was estimated by the court that the State of Alabama would not gain complete control of its teacher testing program "until the year 2015" (*Allen*, Jan. 5, 2000, p. 23). Only recently has a testing company stepped forward with a proposal for a new Alabama teacher certification test (Rawls, 2000).

Plaintiff Richardson was awarded re-employment, backpay, and various other employment benefits (*Richardson*, 1989, pp. 825-26). Defendants (the State of Alabama and its agencies) in both cases were ordered to pay court costs and attorney fees (*Richardson*, 1989, pp. 825-26). However, even though NES was responsible for the development of the tests, NES was not named as one of the defendants in these cases and was not held liable for any damages (Note 7).

Psychometric and Statistical Background

At this point it is appropriate to discuss some of the psychometric concepts and statistics that are fundamental to any question about test quality. The purpose of this discussion is to illustrate that excruciatingly complex analyses are not necessarily required in order to reveal flaws in a test or individual test items. The first steps in test development simply involve common sense practice combined with sound statistical interpretations. If those first steps are flawed, then no complex psychometric analysis will provide a remedy for the mistakes.

One of the simplest statistics reported in the reliability analysis of a test like the MECT is the "item-test point-biserial correlation." This statistic goes by other names such as the "item-total correlation" and the "item discrimination index." It is called the point-biserial correlation specifically because it represents the relationship between a truly dichotomous variable (i.e., an item scored as either right or wrong) and a continuous variable (i.e., the total test score for a person). A total test score, here, is the simple sum of the number of correctly answered items on a test.

The biserial correlation has a long history of statistical use

(Pearson, 1909). One of its earliest measurement uses was as an item-level index of validity (Thorndike, et al., 1929, p. 129). The "point"-biserial correlation appeared specifically for individual dichotomous items in an item analysis because of concerns over the assumptions implicit in the more general biserial-correlation (Richardson & Stalnaker, 1933). It was again used as a validity index. It subsequently came to acquire diagnostic value and was re-labeled as a discrimination index (Guilford, 1936, p. 426).

The purpose of this statistic is to determine the extent to which an individual item contributes useful information to a total test score. Useful information may be defined as the extent to which variation in the total test scores has spread examinees across a continuum of low scoring persons to high scoring persons. In the present situation, this refers to the extent to which well qualified candidates can be distinguished from less capable candidates.

Generally, the greater the variation in the test scores, the greater the magnitude of a reliability estimate. Reliability may be defined many ways through the body of definitions and assumptions known as Classical Test Theory or CTT (Lord & Novick, 1968). According to CTT, an examinee's observed score (X) is assumed to consist of two independent components, a true score component (T) and an error component (E). One relevant definition of reliability may be expressed as the ratio of true-score variance to observed- score variance. Thus, the closer the ratio is to 1.0, the greater the proportion of observed-score variance that is attributed to true-score variance.

The KR-20 reliability estimate is often reported for achievement tests (Kuder & Richardson, 1937, Eq. 20, p. 158). Although reliability as defined above is necessarily positive, the KR-20 can be negative under certain extraordinary conditions (Dressel, 1940) but typically ranges from 0 to +1. Nevertheless, the higher the value, the more "internally consistent" the items on a test. The magnitude of the KR-20, however, is affected by the direction and magnitude of the point-biserial correlations. Specifically, total test score reliability is decreased by the inclusion of items with near-zero point-biserial correlations and is worsened further by the inclusion of items with negative point-biserial correlations. This is because each additional faulty item increases the error variance in the scores at a faster rate than the increase in true-score variance.

Technically, the point-biserial correlation represents the magnitude and direction of the relationship between the set of incorrect (scored as "0") and correct (scored as "1") responses to an individual item and the set of total test scores for a given group of examinees. In other words, it is a variation of the common Pearson product-moment correlation (Lord & Novick, 1968, p. 341). It can range in magnitude from zero to . An estimate near zero is a poorly discriminating item that contributes no useful information. An estimate of +1 would indicate a perfectly discriminating item in the sense that no other items are necessary on the test for differentiating between high scoring and low scoring persons. A value of 1.0 is never attained in practice nor is it sought (Loevinger, 1954). Negative estimates are addressed below.

Ideally the test item point-biserial correlation should be moderately positive. Although various authors differ on what precisely constitutes "moderately positive", a long-standing general rule of thumb among experts is that a correlation of .20 is the minimum to be considered satisfactory (Nunnally, 1967, p. 242; Donlon, 1984, p. 48) (Note 8). There is, however, no disagreement among psychometricians on the direction of the relationship—it has to be positive.

The direction of the correlation is critical. A positive correlation means that examinees who got an item right also tended to score above the mean total test score and those who got the item wrong tended to score below the mean total test score. This is intuitively reasonable and is an intended psychometric outcome. Such an item is accepted as a good "discriminator" because it differentiates between high and low scoring examinees. This is one of the fundamental objectives of classical test theory, the theory underlying the development and use of the MECT.

A negative point-biserial correlation, however, occurs when examinees who got an item correct tended to score below the mean total test score while those who got the item wrong tended to score above the mean total test score. This situation is contrary to all standard test practice and is not an intended psychometric outcome (Angoff, 1971, p. 27). A negative point-biserial correlation for an item can occur because of a variety of problems (Crocker & Algina , 1986). These include:

1. chance response patterns due to a very small sample of people having been tested,
2. no correct answers to an item,
3. multiple correct answers to an item,
4. the item was written in such a way that "high ability" persons read more into the item than was intended and thus chose an unintended distracter while the "low ability" people were not distracted by a subtlety in the item and answered it as intended,
5. the item had nothing to do with the topic being tested, or
6. the item was mis-keyed, that is, a wrong answer was mistakenly keyed as the correct one on the scoring key.

When an item yields a negative point-biserial correlation, the test developer is obligated to remove the item from the test so that it does not enter into the total test score calculations. In fact, the typical commercial testing situation is one where the test contractor administers the test in at least one field trial, discovers problematic items, either fixes the problems or discards the items entirely, and then readministers the test prior to making the test fully operational. The presence of a flawed item on a high-stakes examination can never be defended psychometrically.

One additional point must be made. The point-biserial correlation can be computed two ways. The first way is to correlate the set of 0/1 (incorrect/correct) responses with the total scores as described above. In this way of computing the statistic, the item for which the correlation is being computed contributes variance to the total score, hence, the correlation is necessarily magnified. That is, the statistical estimate of the extent to which an item is internally consistent with the other items "tends to be inflated" (Guilford, 1954, p.439).

The second way in which the correlation may be computed is to compute it between the 0/1 responses on an item and the total scores for everyone but with the responses to that particular item removed from the total score (Henrysson, 1963). This is called the "corrected point-biserial correlation." It is a more accurate estimate of the extent to which an individual item is correlated to all the other items. It is easily calculated and reported by most statistical software packages used to perform reliability analyses (e.g., SPSS's Reliability procedure).

Various concerns have been raised over the interpretation of the point-biserial correlation because the magnitude of the coefficient is affected by the difficulty of the item. The fact is, however, that all the various discrimination indices are highly positively correlated (Nunnally, 1936; Crocker & Algina, 1986). Furthermore, even though the magnitude of the point-biserial correlation tends to be less than the biserial-correlation, all writers agree on the interpretation of negative discriminations. "No test item, regardless of its intended purpose, is useful if it yields a negative discrimination index"(Ebel & Frisbie, 1991, p. 237). Such an item "lowers test reliability and, no doubt, validity as well" (Hopkins, 1998, p. 261). Furthermore, "on subsequent versions of the test, these items [with negative point-biserial correlations] should be revised or eliminated (Hopkins, 1998, p. 259).

NES AND THE MECT

The 1999 MECT Technical Report

In July 1999 NES released their five volume *Technical Report* on the Massachusetts Educator Certification Tests. Volume I describes the test design, item development description, and psychometric results. Volume II describes the subject matter knowledge and test objectives. Volume III consists of "correlation matrices by test field." Volume IV consists of various content validation materials and reports. Volume V consists of pilot material, bias review material, and qualifying score material. The report was immediately hailed by Massachusetts Commissioner of Education David P. Driscoll: "I have said all along that I stand by the reliability and validity of the tests, and this report supports it." (Massachusetts Department of Education, 1999).

Field Trial

Technical Report Volume I contains the psychometric results for the first four administrations of the MECT (April, July, and October 1998, and January 1999). It does not, however, contain any results from a full-scale field trial, nor are any "pilot" test results reported (Note 9). There is no information on how may different items were tested, where the items came from, how many items were revised or rejected, what the revisions were to any revised items, or what the psychometric item-level results were. In fact, there is no field trial evidence in support of the initial inclusion of any of the individual items on the operational exams because *there was no field trial*.

Interestingly, the Department of Education released a brochure in January 1998 stating that the first two test administrations would not count for certification—implying that the tests would serve as a field trial. Chairman of the Board of Education John Silber, however, declared in March 1998 that the public had been misinformed and that the first two tests would indeed count for certification. This policy reversal was unfortunate because of the confusion and anxiety it created among the first group of examinees and because it prevented the gathering of statistical results that could have improved the quality of the test.

NES had considered a field trial of their teacher test in Alabama but did not conduct one and assumedly came to regret that decision. In *Allen v Alabam* they argued, "As the evidence will show, there was no need to conduct a separate large-scale field tryout in this case, since the first test administration served that purpose" (*Allen*, Defendants' Pre-Trial Memorandum, 1986, p. 113). That decision was unwise because it directly affected the implementation and validity of their procedures. For example, "The court has no doubt that, after the results from the first administration of those 35 examinations were tallied, the test developer knew that its cut-score procedures had failed" (*Richardson*, 1989, p. 823). In fact, the original settlement in *Allen v Alabama* stipulated that in any new operational examination, the items "shall be field tested using a large scale field test" (*Allen*, Consent Decree, Oct. 25, 1985, p. 3).

The first two administrations of the MECT would have served an important purpose as a full-scale field trial for the new tests, thus avoiding the mistake made in Alabama. However, that opportunity to detect and correct problems in administration, scoring, and interpretation was lost. The impact of the lack of a field trial is further magnified when it is noted that the time period between when NES was awarded the Massachusetts contract (October 1997) and when the first tests were administered (April 1998) was even smaller than the time period NES had to develop the tests in Alabama—a time frame that the court referred to as "quite short" (*Richardson*, 1989, p. 817). Furthermore, even though NES may have drawn many of the MECT items from existing test item banks, items written and used elsewhere still must be field tested on each new population of teacher candidates.

Point-biserial correlations

In the NES *Technical Report* Volume I, Chapter 8, p. 140, there is a description of when an item is flagged for further scrutiny. One of the conditions is when an item displays an "item-to-test point-biserial correlation less than 0.10 (if the percent of examinees who selected the correct response is less than 50)". After such an item is found, "The accuracy of each flagged item is reverified before examinees are scored." The *Technical Report*, however, does not report or provide the percent of persons who selected the correct response on each item. Nor is there an explanation of what the reverification process consisted of, nor of how many items were flagged, nor what was subsequently modified on flagged items. Thus, there is no way to determine the extent to which NES actually followed its own stated guidelines and procedures in the development of the MECT. The relevance of what NES states as their review procedures and what they actually performed is that in Alabama, under the topic of content validity, it was argued by the defense that items rated as "content invalid" were revised by NES and that these "revisions were approved by Alabama panelists before they appeared on a test." The court, however, found that "no such process occurred" (*Richardson*, 1989, p. 822).

The following table summarizes the point-biserial estimates reported for the MECT. Note that these are not the results prior to NES conducting the item review process. These are the results for the "scorable items" *after* the NES review.

Table 1
Problematic Point Biserial Correlations from the 1999 MECT Technical Report

Date	Number tested	N of M/C Items	Items with point biserials <=0.20					% of total items
			<.00	.00-.05	.06-.10	.11-.15	.16-.20	
Apr-98	4891	315	1	7	15	24	46	29.5%
Jul-98	5716	443	0	2	14	17	39	16.3%
Oct-98	5286	379	2	5	10	15	32	16.9%
Jan-99	9471	507	1	4	14	35	49	20.3%
	25,364	1,644	4	18	53	91	166	332/1644 = 20.2%

Test	Number tested	N of M/C Items	Items with point biserials <=0.20					% of total items
			<.00	.00-.05	.06-.10	.11-.15	.16-.20	
Writing	9750	92	0	0	0	1	1	2.2%
Reading	9455	144	0	0	1	1	6	5.6%
Early Childhood	936	256	0	3	18	30	46	37.9%
Elementary	3125	256	0	2	0	3	27	12.5%
Social Studies	259	128	1	0	1	6	14	17.2%
History	108	64	0	0	2	6	5	20.3%
English	695	256	0	3	11	12	29	21.5%
Mathematics	345	192	1	0	4	4	7	8.3%
Special Needs	691	256	2	10	16	28	31	34.0%
		1,644	4	18	53	91	166	

Source: Massachusetts Educator Certification Tests: Technical Report, 1999

A number of observations may be made from the information in this table. First, of the 1644 total number of items administered over the first four dates, 332 items (20.19%) had point-biserial correlations that are lower than the industry minimum standard criterion of .20. That is a huge percent of poorly performing items for a high-stakes examination. Second, while there are relatively few suspect items on the Reading and Writing tests, there are large numbers of items with poor statistics on many of the subject matter tests. The Early Childhood, English, and Special Needs tests, in particular, consisted of extraordinarily large percentages of poorly performing items (37.9%, 21.5%, and 34%, respectively). Overall, of the 332 items with low point-biserials, 322 (97%) occurred on the subject matter tests. On the face of it, the results for the subject matter tests are terrible. There is, unfortunately, no authoritative source in the literature (including the *Standards*) that tells us unequivocally whether or not this overall 20.19% of poorly performing items on a licensure examination with high-stakes consequences is acceptable, not acceptable, or even terrible. Given the steps that NES claims were followed in selecting items from existing item banks and in writing new items, there simply should not be this many technically poor items on these tests.

Reliability

In Volume I, Chapter 9, p. 188 of the Technical Report, the following statement appears. "It is further generally agreed that reliability estimates lower than .70 may call for the exercise of considerable caution." The practical significance of this statement lies in the fact that when reliability is less than .70, it means that at least 30% of the variance in an examinee's test score is attributable to something other than the subject matter that is being tested. In other words, an examinee's test score consists of less than 70% true-score variance and more than 30% error variance. This ratio of true-score variance to error-variance is not desirable in high-stakes examinations (Haney, et al., 1999). Nearly 40 years ago, Nunnally went so far as to describe as "frightening" the extent to which measurement error is present in high-stakes examinations even with reliability estimates of .90 (1967, p. 226).

NES, however, suggests that their reported item statistics and reliability estimates should not greatly influence one's judgment about the overall quality of the tests because the multiple-choice items make up only part of the exam format (NES, 1999, p. 189). The problem with that argument, as noted by Judge Thompson in *Richardson* (1989, pp. 824-25), is that small errors do accumulate and can invalidate the use for which the test was developed. This issue of simply dismissing troubling statistics as inconsequential is particularly ironic when the MECT has been described by the non-profit Education Trust as "the best [teacher test] in the country" (Daley, Vigue & Zernike, 1999).

The Special Needs test deserves closer attention because it had problems at each reported administration.

1. The sample sizes for the tests were 131, 206, 154, and 200, respectively. Based on NES's own criteria (NES, 1999, p. 187), these sample sizes are sufficient for the generation of statistical estimates that would be relatively unaffected by sampling error.
2. The KR-20 reliability coefficients for the four administrations were .67, .76, .76, and .74, respectively. These are minimally tolerable for the last three administrations. The reliability is not acceptable, however, for the first administration. This means that people were denied certification in Special Needs based on their performance on a test that was deficient even by NES's own guidelines.
3. For the April 1998 administration eleven Special Needs items had point-biserials of .10 or less (again, one of NES's stated criterion for "flagging" an item). For the July 1998 administration it was five items, for October 1998 it was four items, and for January 1999 it was eight items. In fact, in two of the administrations there was an item with a negative point-biserial. (Given the previous discussion about the way the point-biserials were likely to have been calculated (uncorrected), the frequency of negative point-biserials would likely increase if the corrected coefficients had been reported.) Given that there is no specific information about flagging, deleting or replacing items, it is possible that these same faulty items were, and continue to be, carried over from one administration to the next.

The Linkage between Alabama and Massachusetts: A *modus operandi*

At this point the reasonable reader might ask why I am expending so much effort upon what appears to be a relatively minor problem—some items had negative point- biserial correlations. NES, for example, would likely call this analysis "item-bashing", as this type of analysis was referred to in Alabama. The significance of these findings lies in the apparent connection between NES's work in Alabama and their present work on the MECT in Massachusetts.

In Alabama, defendants claimed that

> Before any item was allowed to contribute to a candidate's score, and before the final 100 scorable items were selected, the item statistics for all the items of the test were reviewed and any items identified as questionable were checked for content and a decision was made about each such item (*Allen*, Defendants' Pre-Trial Memorandum, 1986, pp. 113-14).

In fact, in Alabama there were negative point-biserial correlations in the original reliability reports generated by NES (their own documents reported negative point-biserial correlations as large as -0.70) and those negative point- biserial correlations for the same scorable items remained after multiple administrations of the examinations. Simply taking out the worst 20 items in each test did not remove all the faulty items since each exam had to have 100 scorable items. As seen above in Table 1, the MECT has statistically flawed items on many tests, these items have been there since the first administration, and they may be the same items still being used in current administrations.

In Alabama, the negative point-biserial correlations led to the discovery of items for which there was no correct answer. Also discovered were items for which there were multiple correct answers and there were items for objectives that had been rated "not as job related." Additionally, items were found to have been mis-keyed on the item analysis scoring forms. Furthermore, those flawed items existed unchanged for the first eight administrations of the tests. They were not revised, deleted, or changed to "experimental" non-scorable status until the ninth administration--one month after the plaintiffs' team agreed to take the case. Defendants argued that "problems with the testing instrument—such as mis-keyed answers" were simply one component of many that is taken into account by the "error of measurement" (*Allen*, Defendants' Pre-Trial Memorandum, 1986, pp. 108- 113). (Note 10)

As noted earlier, poor item statistics may result for many reasons. Of those reasons the only acceptable one is that they may be due to sampling error (chance). That explanation is unlikely with respect to the MECT, however, because the sample sizes are sufficiently large, and the pattern of faulty item statistics persists over time. The extent to which flawed items may exist in the Massachusetts tests can only be determined by release of the student-level item response data and the content of the actual items, something that has not been done to date. Furthermore, such a release of additional technical information, or item response data, or item content is highly unlikely. (Note 11) In Alabama, the statistical results and in-house documents were not produced by NES until the plaintiffs seriously discussed contempt of court actions against NES personnel. Consequently, there is little reason to expect that NES will voluntarily release MECT data or results not explicitly covered in their original confidential contract.

In Alabama there were no independent testing experts appointed or contracted to monitor the test developer's work. This fact led the court to conclude that "The developer's work product was accepted by the state largely on the basis of faith" (*Richardson*, 1989, p. 817). In Massachusetts the original MECT contract called for the contractor to recommend a technical review committee of nationally recognized experts who were external to their organization (MDOE, 1997, Task 2.14.i, p. 11). The committee was to review the test items, test administration, and scoring procedures for validity and reliability and was to report its findings to the Department of Education. NES did not form such an independent technical advisory committee for the MECT nor has a formal independent review of the MECT been undertaken by anyone else.

It is not in the short-term business interests of a testing company to conduct disconfirming studies on the technical quality of their commercial product. The MECT is, of course, a product that NES markets as an example of what they can build for other states who might be interested in certification examinations. It is, however, in the best interests of a state for such studies to be conducted. For example, the Commonwealth of Massachusetts has a statutory responsibility to "protect the health, safety and welfare of citizens" who seek services from licensed professionals (NES, 1999, p. 16). In the present situation "citizens" are defined by the Board of Education as "the children in our schools" (MDOE, Special Meeting Minutes, 1998). What has apparently been lost in all of this is the fact that prospective educators are "citizens" and deserve protection too--protection from a faulty product that can damage the profession of teaching and can alter drastically the career paths of individuals. Educators and the public at large deserve the highest quality certification examinations that the industry is capable of providing. There is ample evidence that the MECT may not be such an examination.

Conclusion

A technical review of the psychometric characteristics of the MECT has been called for in this journal (Haney et al. 1999; Wainer, 1999). The year 2000 and 2001 budgets passed by the Legislature of the

Commonwealth also called for such an independent audit of the MECT. Those budget provisions, however, were vetoed by Governor Cellucci, and the legislature failed to override the vetoes. Until an independent review committee with full investigative authority is convened by the Commonwealth, the only technical material publicly available for independent analysis is the *1999 MECT Technical Report* generated by NES (NES, 1999). (Note 12) One of the important points made by Haney et al, (1999) was that the Massachusetts Department of Education is not the appropriate agency for conducting such a review. Part of my point here is that the only review of the MECT the Commonwealth may ever see is the one prepared by NES of its own test. Such a review clearly raises a concern over conflict-of-interest (Madaus, 1990; Downing & Haladyna, 1996).

Given the national interest in "higher standards" for achievement and assessment, it must be recognized that there are no "gold" standards by which a testing program such as the MECT can be evaluated (Haney & Madaus, 1990; Haney, 1996). This is ironic given how technically sophisticated the testing profession has become. Consequently, without "gold" standards to define test development practice, there are no legislated penalties for faulty products (tests) and there is no enforced protection for the public. Testing companies may lose business if the details of shoddy practice are made known and the public may appeal to the judicial system for damages. But the opportunity for a test taker simply to raise a question about a test that can shape his or her career and to have that question taken seriously by an impartial panel should be the right of every test-taking citizen. (Note 13)

Contrary to former Chairman John Silber's statement to the Massachusetts Board of Education, "there is nothing wrong with this test" (Minutes of the Board, Nov. 11, 1998) and the statement by the chief of staff for the MDOE, Alan Safran, "[the test]does not show who will become a great teacher, but it does reliably and validly rule out those who would not" (Associated Press, 1998), there is ample evidence that there may be significant psychometric problems with the MECT. These problems, in turn, have significant practical ramifications for certification candidates and the institutions responsible for their training.

Is the MECT sound enough to support assertions that the candidates are "idiots"? No. Is there evidence that poor performance may, in part, reflect a flawed test containing defective items? Yes. Should the Massachusetts Commissioner of Education independently follow through on the twice-rejected Senate bill to "select a panel of three experts from out-of-state from a list of nationally qualified experts in educational and employment testing, provided by the National Research Council of the National Academy of Sciences, to perform a study of the validity and reliability of the Massachusetts educator certification test as used in the certification of new teachers and as used in the elimination of certification approval of teacher preparation programs and institutions to endorse candidates for teacher certification?" (Massachusetts, 1999, Section 326. (S191K)).

Absolutely. Should such a panel serve as a blueprint for the formation of a standing national organization for test review and consumer protection? Yes.

As we enter the 21st century, high stakes tests are becoming increasingly powerful determinants of students' and teachers' lives and life chances. Title II of the 1998 Higher Education Act, in particular, has encouraged a kind of de facto national program of teacher testing. Given the extraordinarily high stakes of these tests, the personal and institutional consequences of poorly designed teacher tests have become too great simply to allow test developers to serve as their own (and lone) quality control and their own (and often non-existent) dispute resolution boards.

Now is the time for the community of professional educators and psychometricians to take a stand and demand that test developers be held accountable for their products in the test marketplace. What this would require at the very least are (1) a mechanism for an independent external audit of the technical characteristics of any test used for high stakes decisions, and (2) a mechanism for the resolution of disputed scores, results, and cases.

Only then will taxpayers, educators, and test candidates have confidence that teacher tests are actually providing the information intended by legislative actions to raise educational standards and enhance teacher quality. Title II legislation certainly did not cause the high stakes test Juggernaut that is rolling through all aspects of educational reform in the U.S. and elsewhere. With mandatory teacher test reporting now tied to federal funding, however, Title II legislation certainly has added to the size, weight, and power of the test Juggernaut and strengthened its hold on reform. For this reason, federal policy makers are now responsible for providing legislative assurances that the public will be protected from the shoddy craftsmanship of some tests and some testing companies and that there will be remedies in place to right the mistakes that result from negligence. This article ends with a call to action. Policy makers must now incorporate into the federal legislation that requires state teacher test reporting new concomitant requirements for the establishment of independent audits and dispute resolution boards.

Notes

I wish to thank Marilyn Cochran-Smith, Walt Haney, Joseph Herlihy, Craig Kowalski, George Madaus, and Diana Pullin for their advice and editorial comments.

1. The class consisted of "all black persons who have been or will be denied any level teaching certificate because of their failure to pass the tests by the Alabama Initial Teacher Certification Testing Program." (Order On Pretrial Hearing, 1984).
2. This specific wording does not appear until the Amended Consent Decree of Jan. 5, 2000.

3. Among other things, conditions were set on the development of new tests, an independent monitoring and oversight panel was established, grade point averages were ordered to be considered in the certification process, and defendants would pay compensatory damages to the plaintiffs and plaintiffs' attorneys' fees and costs (Consent Decree, 1985).
4. That decision has been upheld numerous times since. The latest Amended Consent Decree was approved on January 5, 2000 (*Allen*, Jan.5 , 2000).
5. George Madaus, Joseph Pedulla, John Poggio, Lloyd Bond, Ayres D'Costa, Larry Ludlow.
6. "Failure tables" consisted of an applicant's name, their raw scores on the exams, the exam cut-scores, their actual responses to suspect items, and their recomputed raw scores if they should have been credited with a correct response to a suspect item. Examinees were identified in court who had failed an examination by one point (i.e., missed the cut- score by one item) but had actually responded correctly to a miskeyed item. For example, on the fifth administration of the Elementary Education exam there were six people who should have been scored correct on scorable item #43 (the so-called "carrot" item) but were not. Their total scores were 72. The cut-score was 73. These individuals should have passed the examination. There was even a candidate who took an exam multiple times and failed but who should have passed on each occasion.
7. The standard contract for test development will include some specification of indemnification. In the case of a state agency like the MDOE, the Request For Responses will typically specify protection for the state, holding the contractor responsible for damages (MDOE, 1997, V. (G), 1, p.17). Contractors, understandably, are reluctant to enter into such an agreement and have been successful in striking this language from the contract.
8. The rationale is that .20 is the minimum correlation required to achieve statistical significance at alpha=.05 for a sample size of 100. This is because .20 is twice the standard error (based on a sample of 100) needed to differ significantly from a correlation of zero.
9. The difference between piloting test items, as NES did, and conducting a field-trial is that the field-trial simulates the actual operational test-taking conditions. Its value is that problems can be detected that are otherwise difficult to uncover. For example, non-standardized testing conditions created numerous sources of measurement error on the first administration of the MECT (Haney et al, 1999).
10. This interpretation of measurement error goes considerably beyond conventional practice where "Errors of measurement are generally viewed as random and unpredictable." (*Standards*, 1999, p. 26). A miskeyed answer key is not a random error. It is a mistake and its effect is felt greatest by those near the cut-score. Although false-positive passes may benefit from the mistake, it is

the false-negative fails who suffer and, as a consequence, seek a legal remedy.

11. To date the MDOE has routinely ignored questions requesting technical information, e.g. how many items originally came from item banks, who developed the item banks, how many items have been replaced, what are the reliabilities of new items, what are the technical characteristics of the present tests, will the Technical Report be updated, what "disparate impact" analyses have been conducted?
12. From the start of testing to the present time individual IHE's have not been able to initiate any systematic analysis of their own student summary scores, let alone any statewide reliability and validity analyses. The primary reason for this paucity of within- and across- institution analysis is because NES only provides IHEs with student summary scores printed on paper—no electronic medium is provided for accessing and using one's own institutional data. Thus, each IHE faces the formidable task of hand-entering each set of scores for each student for each test date. This results in a unique and incompatible database for each of the Commonwealth's IHEs.
13. I assert that the right to question any aspect of a high-stakes examination should take precedence over the waiver required when one takes the MECT: "I waive rights to all further claims, specifically including, but not limited to, claims for negligence arising out of any acts or omissions of the Massachusetts Department of Education and the Contractor for the Massachusetts Educator Certification Tests (including their respective employees, agents, and contractors)" (MDOE, 2001, p. 28).

References

Angoff, W. (ed.). (1971). *The College Board Admissions Testing Program: A Technical Report on Research and Development Activities Relating to the Scholastic Aptitude Test and Achievement Tests*. NY: College Entrance Examination Board.

Allen v. Alabama State Board of Education, 612 F. Supp. 1046 (M.D. Ala. 1985).

Allen v. Alabama State Board of Education, 636 F. Supp. 64 (M.D. Ala. Feb. 5, 1986).

Allen v. Alabama State Board of Education, 816 F. 2d 575 (11th Cir. April 22, 1987).

Allen v. Alabama State Board of Education, 976 F. Supp. 1410 (M.D. Ala. Sept. 8, 1997).

Allen v. Alabama State Board of Education, 190 F.R.D. 602 (M.D. Ala. Jan. 5, 2000).

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington, D.C.: American Educational Research Association.

Associated Press Archives, (October 4, 1998). *State Administers Teacher Certification Test Amid Ongoing Complaints*.

Baldus, D.C. & Cole, J.W.L. (1980). *Statistical Proof of Discrimination*. NY: McGraw-Hill.

Cochran-Smith, M. (in press). The outcomes question in teacher education. *Teaching and Teacher Education*.

Cochran-Smith, M. & Dudley-Marling, C. (in press). The flunk heard round the world. *Teaching Education*.

Consent Decree, *Allen v. Alabama State Board of Education*, No. 81-697-N (M.D. Ala. Oct. 25, 1985).

Crocker, L. & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. NY: Holt, Rinehart and Winston.

Daley, B. (1999). "Teacher exam authors put to the test". Boston Globe, 10/7/98, B3.

Daley, B.; Vigue, D.I. & Zernike, K. (1999) "Survey says Massachusetts Teacher Test is best in US". *Boston Globe*, 6/22/99, B02.

Defendant's Pre-trial Memorandum, *Allen v. Alabama State Board of Education*, No. 81-697-N (M.D. Ala. May 1, 1986).

Donlon, T. (ed.) (1984). *The College Board Technical Handbook for the Scholastic Aptitude Test and Achievement Tests*. NY: College Entrance Examination Board.

Downing, S. & Haladyna, A. (1996). A model for evaluating high stakes testing programs: Why the fox should not guard the chicken coop. *Educational Measurement: Issues and Practice, 15:1*, pp.5-12.

Dressel, P.L. (1940). Some remarks on the Kuder-Richardson reliability coefficient. *Psychometrika*, 5, 305-310.

Ebel, R.L. & Frisbie, D.A. (1991) (5th ed.). *Essentials of Educational*

Measurement. NJ: Prentice Hall.

Guilford, J.P. (1936) (1st ed.). *Psychometric Methods*. NY: McGraw-Hill.

Guilford, J.P. (1954) (2nd ed.). *Psychometric Methods*. NY: McGraw-Hill.

Haney, W., & Madaus, G. F. (1990). Evolution of Ethical and Technical standards. In R.K. Hamilton, & J. N. Zaal (Eds.), *Advances in Educational and Psychological Testing* (pp.395-425).

Haney, W.M., Madaus, G.F. & Lyons, R. (1993). *The Fractured Marketplace for Standardized Testing*. Boston: Kluwer.

Haney, W. (1996). Standards, Schmandards: The need for bringing test standards to bear on assessment practice. Paper presented at the annual meeting of the American Educational Research association annual meeting. NY: NY.

Haney, W., Fowler, C., Wheelock, A, Bebell, D. & Malec, N. (1999). Less truth than error?: An independent study of the Massachusetts Teacher Tests. *Education Policy Analysis Archives, 7(4)*. Available online at http://epaa.asu.edu/epaa/v7n4/.

Henrysson, S. (1963). Correction for item-total correlations in item analysis. *Psychometrika, 28*, 211-218.

Hopkins, K.D. (1998) (8th ed.). *Educational and Psychological Measurement and Evaluation*. Boston: Allyn and Bacon.

Kuder, G.F. & Richardson, M.W. (1937). The theory of the estimation of test reliability. *Psychometrika, 2*, 151-160.

Loevinger, J. (1954). The attenuation paradox in test theory. *Psychological Bulletin, 51*, 493-504.

Lord, F.M. & Novick, M.R. (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.

Madaus, G. (May 19-20, 1986). Testimony in *Allen v Alabama* (81-697-N).

Madaus, G. (1990). Legal and professional issues in teacher certification testing: A psychometric snark hunt. In J.V. Mitchell, S. Wise, & B. Plake (Ed.), *Assessment of teaching: Purposes, practices, and implications for the profession*. (pp. 209-260). Hillside, NJ: Lawrence Erlbaum Associates..

Massachusetts. (1999). *FY 2000-2001 Budget*.

Massachusetts Department of Education (February 24, 1997). Massachusetts Teacher Certification Tests of Communication and Literacy Skills and Subject Matter Knowledge: Request for Responses (RFR).

Massachusetts Department of Education (July 1, 1998). Board of Education Special Meeting Minutes. *http://www.doe.mass.edu/boe/minutes/98/min07 0198.html.*

Massachusetts Department of Education (July 27, 1999). Department of Education Press Release. *http://www.doe.mass.edu/news/archive99/pr072 799.html.*

Massachusetts Department of Education (November 28, 2000). Board of Education Regular Meeting Minutes. *http://www.doe.mass.edu/boe/minutes/00/1128r eg.pdf.*

Massachusetts Department of Education (February 16, 2001). Massachusetts Educator Certification Tests: Registration Bulletin. *http://www.doe.mass.edu/teachertest/bulletin 00/00bulletin.pdf*

Melnick, S. & Pullin, D. (1999, April). *Teacher education & testing in Massachusetts: The issues, the facts, and conclusions for institutions of higher education*. Boston: Association of Independent Colleges and Universities of Massachusetts.

Millman, J. (June 17, 1986). Testimony in *Allen v Alabama* (81-697-N).

National Board on Educational Testing & Public Policy. (2000). *Policy statement*. Chestnut Hill, MA: Lynch School of Education, Boston College.

National Commission on Testing and Public Policy. (1990). *From Gatekeeper to Gateway: Transforming Testing in America*. Chestnut Hill, MA: Lynch School of Education, Boston College.

National Evaluation Systems. (1999). *Massachusetts Educator Certification Tests Technical Report*. Amherst, MA: National Evaluation Systems.

Nunnally, J. (1967). *Psychometric Theory*. NY: McGraw- Hill.

Order On Pretrial Hearing, *Allen v. Alabama State Board of Education*, No. 81-697-N (M.D. Ala. Dec. 19, 1984).

Pearson, K. (1909). On a new method of determining correlation between a measured character A and a character B, of which only the percentage of cases wherein B exceeds or falls short of a given intensity is recorded for each grade of A. *Biometrika, Vol. VII.*

Pressley, D.S. (1998). "Dumb struck: Finneran slams 'idiots' who failed teacher tests." *Boston Herald*, 6/26/98 pp. 1,28.

Rawls, P. (2000). "ACT may design test for Alabama's future teachers." *The Associated Press*, 7/11/00

Richardson v. Lamar County Board of Education, 729 F. Supp. 806. (M.D. Ala 1989) *aff'd*, 935 F. 2d 1240 (11th Cir. 1991).

Richardson, M.W. & Stalnaker, J.M. (1933). A note on the use of bi-serial r in test research. *Journal of General Psychology, 8*, 463-465.

Thorndike, E.L., Bregman, M.V., Cobb, Woodyard, E. et al., (1929) *The Measurement of Intelligence*. NY: Teachers College, Columbia University.

U.S. Department of Education, National Center for Education Statistics. *Reference and Reporting Guide for Preparing State and Institutional Reports on the Quality of Teacher Preparation: Title II, Higher Education Act*, NCES 2000- 089. Washington, DC: 2000.

Wainer, H. (1999). Some comments on the Ad Hoc Committee's critique of the Massachusetts Teacher Tests. *Education Policy Analysis Archives, 7(5).* Available online at http://epaa.asu.edu/epaa/v7n5.html.

Walden, J.C. & Deaton, W.L. (1988). Alabama's teacher certification test fails. 42 *Ed. Law Rep.1*

About the Author

Larry H. Ludow
Associate Professor
Boston College
Lynch School of Education
Educational Research, Measurement, and Evaluation Department

Email: Ludlow@bc.edu

Larry Ludlow is an Associate Professor in the Lynch School of Education at Boston College. He teaches courses in research methods, statistics, and psychometrics. His research interests include teacher testing, faculty evaluations, applied psychometrics, and the history of statistics.

Copyright 2001 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Sherman Dorn
University of South Florida

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
hmwkhelp@scott.net

Thomas F. Green
Syracuse University

Alison I. Griffith
York University

Arlen Gullickson
Western Michigan University

Ernest R. House
University of Colorado

Aimee Howley
Ohio University

Craig B. Howley
Appalachia Educational Laboratory

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Benjamin Levin
University of Manitoba

Thomas Mauhs-Pugh
Green Mountain College

Dewayne Matthews
Western Interstate Commission for Higher Education

William McInerney
Purdue University

Mary McKeown-Moak
MGT of America (Austin, TX)

Les McLean
University of Toronto

Susan Bobbitt Nolen
University of Washington

Anne L. Pemberton
apembert@pen.k12.va.us

Hugh G. Petrie
SUNY Buffalo

Richard C. Richardson
New York University

Anthony G. Rud Jr.
Purdue University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Jay D. Scribner
University of Texas at Austin

Michael Scriven
scriven@aol.com

Robert E. Stake
University of Illinois—UC

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

U.S. Department of Education Brigham Young University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

other vols. | abstracts | editors | board | submit | comment | subscribe | search

This article has been retrieved 1308 times since March 4, 2001

Education Policy Analysis Archives

Volume 9 Number 7 **March 4, 2001** **ISSN 1068-2341**

A peer-reviewed scholarly journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2001, the **EDUCATION POLICY ANALYSIS ARCHIVES**.
Permission is hereby granted to copy any article
if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Critique of "An Evaluation of the Florida A-Plus Accountability and School Choice Program"

Gregory Camilli
Rutgers University

Katrina Bulkley
Rutgers University

Related article.

Abstract

In 1999, Florida adopted the "A-Plus" accountability system, which included a provision that allowed students in certain low-performing schools to receive school vouchers. In a recently released report, *An Evaluation of the Florida A-Plus Accountability and School Choice Program* (Greene, 2001a), the author argued that early

> evidence from this program strongly implies that the program has led to significant improvement on test scores in schools threatened with vouchers. However, a careful analysis of Greene's findings and the Florida data suggests that these strong effects may be largely due to sample selection, regression to the mean, and problems related to the aggregation of test score results.

One of the most closely watched state reforms in recent years is the use of school vouchers as a part of the accountability system for Florida's public schools. This program is of particular interest because of its strong similarities with proposals put forward by President George W. Bush. As a New York Times article noted, "Gov. Jeb Bush's educational program in Florida has been held up as a model for its combination of aggressive testing of schools' performance, backed by taxpayer-financed vouchers, which his brother President Bush is proposing for the nation as a whole" (Schemo, 2001).

A recently published report purports to show a convincing link between the threat of school vouchers for students in certain low-performing schools in Florida and achievement gains in those schools. *An Evaluation of the Florida A-Plus Accountability and School Choice Program* (Greene, 2001a) documents gains in achievement on the Florida Comprehensive Assessment Test (FCAT) in the areas of reading, mathematics, and writing. (This evaluation will be referred to as *Evaluation of Florida's A-Plus Program*, for short.) These findings, not surprisingly, have received a substantial amount of attention in the popular press (cf. Schemo, 2001; Lopez, 2001; Greene, 2001b). The gains reported are attributed to incentives implemented under Title XVI (section 229.0535 "Authority to enforce school improvement") of the 2000 Florida Statutes:

> It is the intent of the Legislature that all public schools be held accountable for students performing at acceptable levels. A system of school improvement and accountability that assesses student performance by school, identifies schools in which students are not making adequate progress toward state standards, institutes appropriate measures for enforcing improvement, and provides rewards and sanctions based on performance shall be the responsibility of the State Board of Education.

In the A- Plus accountability system, schools are evaluated and assigned one of five grades (A, B, C, D, F) based primarily on FCAT scores, and to a lesser extent, the percent of eligible students tested and dropout rates (Florida Department of Education, 2001). If a school receives two grades of "F" in any four-year period, it becomes eligible for state board action. Contrary to the implication in Greene's title, such action is not limited to school choice; rather, actions may include providing additional resources, implementing a school plan or reorganization, hiring a new principal or staff, and other unspecified remedies designed to improve performance. However, the possibility of public schools losing children to either private schools or higher-performing public schools is clearly the area of most interest and controversy. In the 1999-2000 school year, two Pensacola elementary schools met the eligibility criteria (Note 1), and as a result, lost 53 children to private schools and 85 to other public schools.

Greene argued that his report "shows that the performance of students on academic tests improves when public schools are faced with the prospect that their students will receive vouchers" (p. 2). At the center of his argument is the fact that all 78 schools that received an "F" in 1999 received a higher grade in 2000. His claim that the threat of vouchers was responsible for the improvement of "F" schools (from the 1998-1999 to the 1999-2000 school year) includes several important elements. First, an attempt was made to show the validity of the FCAT by showing a strong correlation to another test (Stanford-9) given in Florida in 2000. Given this evidence, he then proceeded to show the average gains for each school receiving a particular grade. Based on the latter results, it was concluded that:

> The most obvious explanation for these findings is that an accountability system with vouchers as the sanction for repeated failure really motivates schools to improve. (p. 9)

However, Greene also wrote:

> While the evidence presented in the report supports the claims of advocates of an accountability system and advocates of choice and competition in education, the results cannot be considered definitive. (p. 9)

The A-Plus accountability system was duly noted as being relatively new, with the voucher options used in only two schools in the state, and possible—though not likely—manipulation of FCAT scores. It is an additional alternative that Greene mentions, commonly known as *regression to the mean*, that is one main concern of this report. This paper also examines three other issues: (1) sample selection, (2) the combining of gain scores across grade levels, and (3) the use of the school as the unit of analysis. Below, we subsume the latter two items under the category of "aggregation."

The potential policy importance of the findings Greene reports places a heavy burden on his study to demonstrate that the improved scores in schools that had previously received one "F" are in fact meaningful improvements and a result of school changes linked to the threat of vouchers. We argue here that the evidence does not support this conclusion. We show that there may have been some small achievement gains in Florida from 1999-2000, but these effects were vastly overestimated in Greene's analysis. However, even if these modest outcomes withstand further investigation, it is not at all clear that they resulted from the threat of vouchers as opposed to other aspects of the accountability program.

Background

Several recent reforms have similar components to the Florida effort. It is not the purpose of this report to review that literature, but two well-known reforms deserve mention. One of these, which Greene specifically addresses, is the Texas accountability system and its use of the Texas Assessment of Academic Skills (TAAS). Another is the public voucher program in the city of Milwaukee. Comparisons between each of these reforms and the Florida's A-Plus accountability system are limited for a variety of reasons. The accountability system in Texas varies in critical ways from the model in Florida, especially in the use of vouchers as a sanction in the latter state but not the former. Greene did, however, address an important methodological concern (discussed below) that arose in a recent study of the TAAS (Klein, Hamilton, McCaffrey, and Stecher, 2000). In the area of publicly-funded vouchers, students in Milwaukee who met certain income requirements are eligible to receive vouchers allowing them to attend local private schools. Several evaluations have been done of this program (i.e. Witte, 1996; Greene, Peterson and Du, 1998). These evaluations are not comparable to the Florida evaluation because they examined the test scores of individual students who either received vouchers or applied for vouchers but did not receive one; the Greene study focuses on the school impact on test scores of the threat of vouchers, not the actual provision of vouchers.

Summary of the *Evaluation of Florida's A-Plus Program*

In *Evaluation of Florida's A-Plus Program* (Greene, 2001a, Table 2), the main results were obtained by aggregating across grade for school types A, B, C, D, and F. These results are reproduced in Table 1 below.

Table 1
FCAT Reading and Mathematics 1999-2000 Gains from Greene's "An Evaluation of the Florida A-Plus Accountability and School Choice Program"

Grade	Reading	Math	Writing
A	1.90	11.02	.36
B	4.85	9.30	.39
C	4.60	11.81	.45
D	10.62	16.06	.52
F	17.50	25.66	.87

BEST COPY AVAILABLE

To obtain the overall reading and writing gain, gains at the 4th, 8th, and 10th grade levels were pooled, while for mathematics, gains at the 5th, 8th, and 10th grade levels were pooled. School means for standard curriculum students were used to compute gains, not individual student scores. It can be seen that the average gain for "F" schools "are more than twice as large as those experienced in schools with higher state-assigned grades" (Greene, 2001a, p. 6). These gains for "F" schools were then translated into effect sizes for reading (.80), mathematics (1.25), and writing (2.23) (Greene, 2001a, endnotes 12-14). No doubt, as computed, these gains are statistically significant. They are also among the highest gains ever recorded for an educational intervention. Results like these, if true, would be nothing short of miraculous, far outpacing the reported achievement gains in Texas and North Carolina. This may have moved Greene to conclude:

> While one cannot anticipate or rule out all plausible alternative explanations for the findings reported in this study, one should follow the general advice to expect horses when one hears hoof beats, not zebras. The most plausible interpretation of the evidence is that the Florida A-Plus system relies upon a valid system of testing and produces the desired incentives to failing schools to improve their performance. (p. 14)

Critique of the *Evaluation of Florida's A-Plus Program*

Our critique of Greene's evaluation focuses primarily on two problematic issues: aggregation and regression to the mean. We do not examine in detail Greene's validation argument for the FCAT based on its correlations with the Stanford-9 (the latter given in 2000). Greene's correlational analysis was conducted partly in response to concerns raised by Klein and his colleagues (2000) about the validity of the TAAS in Texas. However, it is worth noting that while the two tests have substantial correlations (in the range .85-.95), correlation coefficients computed on aggregate scores typically have much higher values than those computed with student scores. For example, school means on the reading and mathematics sections of the FCAT in 8th grade have a correlation of about .96. This correlation should *not* be interpreted as meaning that the FCAT reading and mathematics tests are statistically indistinguishable, but rather that correlations on aggregate score tend to be much higher than those for individual scores.

Sample Selection

Greene (2001a) used the school means of "standard curriculum" students to obtain school-level gains scores. Here "standard" defines a subset of students who tend to score higher on the FCAT (i.e., it does not include certain types of students with disabilities). An alternative method of choosing a sample is to use the results for all curriculum groups, and these data are available on the Florida Department of Education web pages. While there is nothing intrinsically wrong with using standard curriculum students, for the purposes of evaluation, however, it would seem preferable to look at the potential impact of the A-Plus program on all curriculum groups. Florida administrative statues allow for (or require) nontrivial variation in populations selected for determining school grades (Note 2).

Aggregation

In the analyses below, we disaggregate results by grade. This is useful because overall state gains (Florida Department of Education, 2001) vary by grade as shown in Table 2.

Table 2
FCAT Score Gains from School Year 1998-1999 to 1999-2000

Grade	Reading	Math	Writing
4	5.0	N/A	0.0
5	N/A	11.0	N/A
8	-5.0	7.0	0.0
10	-4.0	3.0	0.1

The data in Table 2 suggest several problems with aggregation across grades. First, the results of a policy implementation may be different at different grades, even if this is not an a priori expectation. Second, in order to fine-tune a successful policy—or weed-out an unsuccessful policy—suitable diagnostic information is critical. Furthermore, a subtle problem arises when mixing the scales of two different instruments given at different grades. How can we be sure that this isn't the old apples- and-oranges problem? To be safe, the best advice is to conduct separate analyses and then to combine them while making explicit the assumptions involved.

A more subtle problem involves the computation of effect size (Hedges, 1985), which is typically taken to be

$$\delta = \frac{x - E[x]}{\sigma}$$

This formula can be read as the difference between an observed value and an expectation divided by the standard deviation. In practice, the expectation $E[x]$ could be a school's average test score for the prior year, and x could be taken as the score for the current year. It is also typical practice to use a measure of student individual variation in the denominator for "sigma" to facilitate a standard interpretation. For example, $\delta = 1$ means that the average student in the "treatment" population scores at the 84th percentile of the "control" population. Likewise, $\delta = 2$ means that the average student in the "treatment" population scores at the 98th percentile of the "control" population. So the interpretation is anchored in individual student achievement.

In contrast, Greene computed effect sizes relative to the standard deviation (SD) of schools, and though this is technically defensible, it must be recognized that such an effect size doesn't have the usual interpretation. In fact, we have estimated that the individual-level standard deviations (SD) are about 70 score points for reading and mathematics, and about .85 for writing—while the school-level SDs are about 20 points for reading and writing, and about .39 point for writing. Thus, an effect size for reading based on the school-level SD would be 350% larger than one based on the individual-level SD. At face value, the effect sizes computed by Greene, ranging from .80 to 2.23, are implausible because many studies have found that especially large educational effects (produced under laboratory conditions) fall into the range of .4 - .7.

But even if Greene's effect sizes are rescaled for comparability, they are still inflated by other factors including regression to the mean (see below) and an inappropriately selected definition of the expectation $E[x]$. In regard to the latter issue, the effect of a treatment is usually defined as the net effect above and beyond average growth (the latter is referred to by statisticians as the *grand mean*). Thus, gain is defined as the net effect above average, and loss as the net effect below average. In this case, the average is the overall state gain; and the deviation from the grand mean represents the unique effect of a particular treatment or intervention. For example, take the average state gain for 4th grade reading in Table 2 of 5 points. If an intervention is defined as positive, it should register as being greater than 5 points since 5 points is what could be expected with no intervention whatsoever. It's not very useful to apply this correction to Greene's Table 1 because the results are aggregated across grades. However, in our analyses below, we build in this correction. We also use the individual-level standard deviation to facilitate the comparability of effect sizes to the general research literature.

Regression to the mean

Campbell & Stanley (1966) in their classic volume *Experimental and Quasi-Experimental Designs for Research* defined the *internal validity* of an experiment as:

> The basic minimum without which any experiment is uninterpretable: Did in fact the experimental treatments make a difference in this specific experimental instance? (p. 5)

In a very simple investigation, there are only two measurements taken: the pretest (O_1) and, after the experimental intervention, the posttest (O_2). Campbell and Stanley (1966) listed five definite weaknesses of this "One-Group Pretest-Posttest Design" and one potential concern which is of central importance to Greene's evaluation: *regression to the mean* or, alternatively, *regression artifacts*. They explained:

> If, for example, in a remediation experiment, students are picked for a special experimental treatment because they do particularly poorly on an achievement test (which becomes for them O_1), then on a subsequent testing using a parallel form or repeating the same test, O_2 for this group will almost surely average higher than O_1. This dependable result is not due to any genuine effect of [the intervention], and test-retest practice effect, etc. It is a rather tautological aspect of the imperfect correlation between O_1 and O_2. (p. 10)

In short, experimental units chosen on the basis of extreme scores tend to drift toward the mean upon posttest: low scores drift upward and high score drift downward. Campbell and Stanley (1966) then gave an extended treatment to this topic because "errors of inference due to overlooking regression effects have been so troublesome in educational research," and "the fundamental insight into their nature is so frequently missed" (p. 10). The regression phenomenon emerged from Francis Galton's studies of inheritance in biology, and this subject provides the most common phrasing of the regression to the mean effect: tall fathers tend to have tall sons, but not as tall on average as the fathers; while short fathers have short sons, but not as short on average as the fathers.

It can be seen in Table 1 for all three FCAT subjects that the trend is for higher achievement schools to gains less and lower achievement schools to gain more. This is a tell-tale sign of statistical regression, that is, scores in the tails of the distribution tend to drift toward the mean. Higher scores drift downward and lower scores drift upward relative to average gains. Greene (2001a) did consider this possibility, but rejected it as a potential explanation, arguing that:

> Regression to the mean is not a likely phenomenon for the exceptional improvement made by the F schools because the scores for those schools were nowhere near the bottom of the scale for possible results. The average F school reading score was 254.70 in 1999, far above the lowest possible score of 100.
>
> Likewise, the average FCAT mathematics and writing scores of

the F schools were 272.5 on a scale of 100-500 and 2.40 on a scale form 1-6, respectively. Greene thus concluded that regression to the mean was not a problem because the scores of the F schools were not at all extreme.

This is an inaccurate notion of regression to the mean because "extremeness" should be evaluated in terms of distance (in standard deviation units) below the overall group mean, rather than relative to the lowest possible score. A good measure of "distance below the mean" can be given in z-score units which are interpreted as "standard deviations below the mean" in the distribution of school means; z-scores of –3.00 and lower generally indicate substantial distance below the mean. To check for extremeness, we calculated the z-scores of the lowest performing school in 4th, 8th, and 10th grade reading, and 5th, 8th and 10th grade mathematics. These z-scores ranged from a high of –3.2 to a low of –4.5, indicating a strong likelihood of obtaining a regression artifact in simple difference scores; however, the writing scores tended to be less extreme for the "F" schools.

In North Carolina, it was recognized that "Students who are proficient may grow faster" and "students who score low one year may score higher the next year, partly due to 'regression to the mean'" (Public Schools of North Carolina, 2000, p. 2). Both influences on achievement are explicitly taken into account in the North Carolina system when computing expected growth for schools. As noted by Campbell and Stanley (1966) the incorrect interpretation of regression effects has plagued educational research for decades. To give an example, consider a study by Glass and Robbins (1967) in which the SAT was given to a group of students, and researchers then took the high scorers as the control group and the low scorers as the treatment group. Predictably, the treatment showed a positive effect that disappeared when regression effects were taken into account (Glass & Robbins, 1967)

Methods

Data Sources

The state of Florida has an exceptional policy of granting the public full access to state, district, school level test scores, and other variable such as class size, per pupil expenditures, and the like. These data files containing school means for all curriculum students can be downloaded in the form of Excel spreadsheets at the Florida Department of Education website. For the present analysis, reading and mathematics, and writing FCAT scores at the school level were downloaded for both the 1998-1999 and 1999-2000 school years. Department staff provided a spreadsheet containing school grades, with district and school identification numbers, for the 1998-1999 school year.

Residual gain score analysis

Since we strongly suspected that the statistics in Table 1 were affected by at least two sources of error (regression to the mean and incorrect definition of net effect), we reanalyzed the data using the technique of *residual gain scores*. Glass and Hopkins (1996) described the context for residual gains:

> Administering parallel forms of the achievement test before [O_1] and after [O_2] instruction, then subtracting the pretest score from the posttest score [O_2 - O_1] for each student produces a measure that is far closer to the researcher's notion of a measurement of an achievement gain. One difficulty remains: Such a posttest-minus-pretest measure, [O_2 - O_1], is contaminated by the regression effect, usually correlate negatively with the pretest scores [O_1] ... A better method to measure gain or change is to predict posttest scores [$\boldsymbol{O}_2$] from pretest scores [O_1] and use the deviation [O_2 - $\boldsymbol{O}_2$] as a measure of gain, above and beyond what is predictable by the pretest alone. (p. 167)

In the present case of the FCAT scores, O_1 is the pretest and O_2 is the posttest. Everything else in the present case is the same as in Glass and Hopkins's recommendation. By using residual gains, two goals are accomplished. First, the regression effect is removed because the predicted score takes into account movement toward the mean. Second, the predicted value takes into account the average state gain; it will lead to unique net (policy) effects for any particular accountability grades.

Results

Average residual gains for the FCAT reading and mathematics tests, disaggregated by grade, are given in Tables 3 (reading), 4 (mathematics), and 5 (writing) below.

Table 3
Average Residual Gains for FCAT Reading

GRADE	GROUP	Mean	N	SD
4	A	1.45	121	8.30
	B	3.23	212	10.26
	C	-.86	694	10.54
	D	-.91	455	13.86
	F	2.35	66	12.96
8	A	.44	73	6.94
	B	1.03	90	7.68
	C	-.06	255	8.19
	D	-1.71	94	10.29
	F	7.26	7	12.84

10	A	.79	8	3.89
	B	2.55	12	4.77
	C	-.18	280	6.99
	D	.62	57	8.82
	F	-5.53	4	11.40

Table 4
Average Residual Gains for FCAT Mathematics

GRADE	GROUP	Mean	N	SD
5	A	4.30	121	7.61
	B	.17	210	8.71
	C	-.05	695	10.83
	D	-1.80	449	13.81
	F	4.36	66	15.13
8	A	.39	73	6.98
	B	.32	90	8.54
	C	-.19	255	9.28
	D	-.75	94	10.97
	F	8.78	7	10.82
10	A	1.88	8	2.47
	B	1.59	12	4.45
	C	-.18	280	6.66
	D	.78	57	8.91
	F	-6.73	4	14.73

In Tables 3 and 4, the largest effects are in the 8th grade, but in terms of standard deviation (SD) units, these effects are small (Note 3). Using the individual student SD of about 70 (versus the school SD of about 23), the effect size for 8th grade reading is $\delta = .10$, and for 8th grade math is about $\delta = .13$. We think it is not worthwhile to persevere on whether these effects are statistically significant because they are relatively small and other sources of possible bias cannot be plausibly ruled out as causes. For example, slight nonlinearities in the regressions might account for the higher effect sizes for the 8th grade F schools. In addition, the average effect for this group of only 7 schools is accompanied by a relatively high standard deviation. This means the overall positive effect is highly variable.

The results for FCAT writing are somewhat different for those in reading and mathematics. It can be seen in Table 5 at the 4th grade level that the average residual gain was .20 point on a scale that ranges from 1-6, and this effect is statistically significant. We estimated the individual-level SD to be about .88 point, and consequently the latter gain translates into an effect size of about .23. The average gains are also positive at 8th and 10th grade, but much smaller. Greene also found an effect for writing, but estimated it to have an effect size of 2.23.

Table 5
Average Residual Gains for FCAT Writing

GRADE	TYPE	Mean	N	SD
4	A	.04	121	.19
	B	.03	212	.22
	C	-.02	694	.22
	D	-.01	454	.24
	F	.20	66	.25
8	A	.05	73	.17
	B	.07	90	.18
	C	.00	255	.17
	D	-.05	94	.21
	F	.11	7	.17
10	A	.11	8	.09
	B	.15	12	.15
	C	.01	279	.23
	D	-.03	57	.22
	F	.10	4	.18

Greene attempted to control for regression effects by comparing higher-scoring "F" schools to lower-scoring "D" schools. "These gains made by the higher-scoring F schools in excess of what were produced by the lower-scoring D schools are what we can reasonably estimate as the effect of the unique motivation that vouchers posed to those schools with the F designation" (p. 8). Using residual scores, we repeated this analysis using 40 schools in each of the above categories aggregated across grade for reading and mathematic (though we don't suggest this as an analytic strategy). The estimates of effect were small and nonsignificant.

Discussion

The A-Plus accountability system in Florida, with its inclusion of school vouchers as one possible repercussion for low-performing schools, is a significant policy shift in the use of high-stakes assessment. Findings from evaluations of this program may thus play an important role in policy making in other states and at the federal level. Unfortunately, the Greene evaluation does not meet the methodological demands for such an evaluation. It is clear that Greene's analysis failed to account for both regression to the mean and obtaining a unique net effect of being labeled an "F" school. Sample selection is a debatable issue, and we have argued in this report that indicators based on all curriculum groups better satisfy the demands of evaluation.

Some have argued that information and research must be central to the improvement of schools:

> Schools that consistently fail to educate poor children should not receive federal dollars—and states should be accountable to Washington for ensuring that this does not happen. Federal programs that can't demonstrate results should themselves be replaced by different strategies. Though innovation and experimentation should always be encouraged, rigorous evaluation is vital and federal funds

> should not flow to activities that do not yield results for children. (Finn, Bruno & Ravitch, 2000)

In reply, we would argue that it's not always easy to demonstrate results given the kinds of data and accountability models that are readily available. As seen in Florida, the accountability model itself may cause some difficulty (Note 4). If schools in the lowest classification "F" improve, and yet this "improvement" is a regression artifact, then teachers and principals and others may seize upon wholly irrelevant events as the causes of this improvement. Likewise, "D" schools that move down to the "F" classification may seize upon wholly irrelevant causes for their demise. While it is true that true "F" schools will tend to bounce up and down, and thus be more likely to become eligible for intervention, it is also true that the accountability system as currently structured may provide them with unreliable signs of their progress (or lack thereof).

Positive results are more helpful if they can be shown (by means of high quality evaluations) to be internally consistent with policy mechanisms that presumably stimulated change. One can learn better from negative outcomes if it can be shown in some detail how the policy levers failed. In other words, learning more about *how* schools made improvements or reasons for slippage is important, as well as is having confidence that the measures of loss or gain are both reliable and valid. Tying accountability to a single (or even a few) achievement outcomes has several downsides: (1) it does not automatically increase our knowledge about why things happened the way they did; (2) the use of statistical models for monitoring policy outcomes is technically demanding and requires obscure policy tools such as adjustments for regression to the mean. Moreover, it is problematic to conflate evaluation and accountability: program evaluation is intrinsically important to the mission of schools and should not be equated with establishing "results" as defined by Washington.

We can agree that hard-nosed evaluation is necessary, but it is useful to expand on what such evaluation activities should include:

Technical considerations. The state of Florida should consider methods that are used elsewhere (e.g., North Carolina) to stabilize the indicators that are used to designate school classifications. Such models use past achievement data to estimate expected growth, and designate exemplary growth in a manner that controls for some statistical artifacts such as regression. Though there are costs associated with a more complex model, the decision to focus accountability on test scores requires more sophisticated statistical apparatus within the accountability model. Moreover, focus on a small set of indicators accompanied by significant sanctions can force schools to employ instructional methods that are optimized for short-term payoffs. Consequently, additional accountability components may be required to monitor for negative consequences such as an increase in the number of remedial classes, focusing on test preparation, curricular materials that are substantially similar to test preparation material, and increases in drop-out rates.

Policy considerations. One of the most important roles of policy evaluation is to inform policymakers not only about whether or not a program is working, but *why* it is having the noted effects. Evaluations that provide little or no information about the mechanisms that have led to reported changes are both less compelling and more subject to criticism. In Florida, there is currently little information about what schools are doing that would lead one to expect that scores would improve. This information is crucial for the future development of the accountability program and might include, for example, an evaluation of capacity within schools identified as needing intervention, or an analysis of how administrative rules are interpreted by local staff. Policy makers should also receive evaluation information regarding the accountability model or system itself as well as behavior that is the object of the model.

In the case of Florida, this report suggests that it is simply not clear whether or not the threat of vouchers is having a positive impact on student test scores. There is some evidence of a small effect at 8th grade in reading and mathematics, and in writing at 4th grade. These findings should be investigated in a more thorough analysis (taking into account, for example, exclusion rates). If these findings withstand further analysis, it would also be important to examine a number of potential causes including resources (e.g., professional development or teaching materials), school intervention plans, staffing changes, and other taken remedies to improve performance. In other words, it is overly simplistic to assume that the voucher threat was the only active agent, or that other causes were contingent on the voucher threat.

Conclusion

We offer an alternative to Greene's generous and simplistic reading of the evidence. At face value, the large gains (as seen in effect sizes of .80, 1.25, and 2.23 for reading, mathematics, and writing) were implausible and should have been submitted to additional methodological scrutiny. Upon such an examination, we have raised serious questions regarding the validity of Greene's empirical results and conclusions. Indeed, one *should* follow the general advice to expect horses when one hears hoof beats, not unicorns.

Notes

1. These two schools were chosen in 1999 for the voucher plan in the first year of the accountability policy implemented in 1999. These schools did not meet the "2-out- of-4" policy, but had received an "F" in 1999 and appeared on a 1998 list of low performing schools (Sandham, 1999).
2. It could be argued that the group of students who were used to determine the school grade might also be the appropriate sample. It appears that "standard curriculum" designates eligible

students. According to the State Board of Education Administrative Rules (6A-1.09981)

> (3)(a) For the purpose of calculating state and district results, the scores of all students enrolled in standard curriculum courses shall be included. This includes the scores of students who are speech impaired, gifted, hospital homebound, and Limited English Proficient (LEP) students who have been in an English for Speakers of Other Languages (ESOL) program for more than two (2) years.

To receive a grade of "D" or higher, schools are required to test at least 90% of their eligible students. There are additional restrictions on student inclusion for determining school grade in 6A-1.09981:

> (3)(b) For the purpose of designating a school's performance grade, only the scores of those students used in calculating state and district results who are enrolled in the second period and the third period full-time equivalent student membership survey as specified in Rule 6A-1.0451, FAC., shall be included.

Because these criteria, fairly applied, may create inconsistencies across schools, the group of all students tested may provide a school average better for the purposes of evaluation. It would also be useful to have the school median and exclusion rates.

3. The frequencies in Tables 3 and 4 differ slightly from the actual number of schools in each category. For example, 5 high schools received grades of "F," yet there are only 4 in our study. In checking this result, we found that the 5th high school was no longer listed in official documents in 1999-2000. Other than this difference, however, our data agree with state data in terms of the numbers of "F" schools for elementary, middle, and high schools.
4. We note that only two of the schools on the 1998 list of critically low performing schools received an "F" in 1999. Likewise none of the 78 schools receiving an "F" in 1999 also received an "F" in 2000; however, only 4 schools received an "F" in 2000.

References

Campbell, D.T., and Stanley, J.C. (1966). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally.

Finn, C. E., Manno, B. V., Jr. & Ravitch, D. (2000). Education 2001: Getting the Job Done A Memorandum to the President-Elect and the 107th Congress. Thomas B. Fordham Foundation, Washington, D.C.: author.

Florida Department of Education. (2001). *FCAT Briefing Book*, Florida Department of Education, Tallahassee: author.

Glass, G. V & Hopkins, K.D. (1996). *Statistical Methods in Education and Psychology (3rd Ed.)*. Boston: Allyn and Bacon.

Glass, G.V & Robbins, M.P. (1967). A critique of experiments on the role of neurological organization in reading performance. *Reading Research Quarterly*, 3, 5-51.

Greene, J. P. (2001a). *An Evaluation of the Florida A-Plus Accountability and School Choice Program*. New York: The Manhattan Institute.

Greene, J. P. (2001b). Bush's School Plan: Why We Know It'll Work. *New York Post* (February 21).

Greene, J. P., Peterson, P. E., & Du, J. (1998). School Choice in Milwaukee: A Randomized Experiment. In P. E. Peterson & B. C. Hassel (Eds.), *Learning from School Choice*. Washington, D.C.: Brookings Institution Press.

Hedges, L.V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Klein, S. P., Hamilton, L. S., McCaffrey, D. F. & Stecher, B. M. (2000). What Do Test Scores in Texas Tell Us? *Education Policy Analysis Archives,* 8 (49).

Lopez, K. J. (2001). Bush Ed Plan Rates an "A." *National Review Online* (February 22).

Public Schools of North Carolina. (2000). Setting Annual Growth Standards: "The Formula." North Carolina Department of Public Instruction, North Carolina: author.

Sandham, J. L. (1999). Schools Hit by Vouchers Fight Back. *Education Week* (September 15).

Schemo, D. J. (2001). Threat of Vouchers Motivates Schools to Improve, Study Says. *New York Times* (February 16).

Witte, J. F. (1996). Who Benefits from the Milwaukee Choice

Program? In B. Fuller & R. F. Elmore (Eds.), *Who Chooses? Who Loses?* New York: Teachers College Press.

About the Authors

Gregory Camilli
Professor
Rutgers Graduate School of Education

Email: camilli@rci.rutgers.edu

Gregory Camilli is Professor of Educational Psychology, at the Rutgers Graduate School of Education and a Senior Researcher in the Center for Educational Policy Analysis. His areas of research interest include psychometric issues in educational policy, meta-analysis, and differential item functioning. Examples of recent publications include "Values and state ratings: An examination of the state-by-state education indicators in quality counts" (*Educational Measurement: Issues and Practice*, 2000), "Application of a method of estimating DIF for polytomous test items" (*Journal of Educational and Behavioral Statistics,* 1999), Standard error in educational programs: A policy analysis perspective" (*Educational Policy Analysis Archives*, 1996), and *Methods for Identifying Biased Items* (Sage, 1994). Camilli has been or is currently a member of the editorial Boards of *Educational Measurement: Issues and Practice*, *Educational Policy Analysis Archives*, and *Education Review*. He is a regular reviewer for *Applied Measurement in Education, Journal of Educational Measurement, Psychometrika*, and *Psychological Methods*, among others. As a member of the Technical Advisory Committee of the New Jersey Basic Skills Assessment Council, he provides expertise on testing and measurement issues to the New Jersey state assessment program.

Katrina Bulkley
Assistant Professor
Rutgers Graduate School of Education

Email:bulkley@rci.rutgers.edu

Katrina Bulkley is an Assistant Professor of Educational Policy at the Rutgers University Graduate School of Education. Much of her work has focused on issues involving school choice and charter schools. Recent articles include, "Charter School Authorizers: A New Governance Mechanism?" in *Educational Policy* (November 1999), and "'New Improved' Mayors Take Over City Schools" (with Michael Kirst) in Phi Delta *Kappan* (March 2000). She is currently working with the Consortium for Policy Research in Education on a literature review of research on charter schools and a study of for-profit management companies and charter schools, and with the Center for Education Policy Analysis, located at Rutgers University, on two studies of the impact of standards, testing and professional development on instructional practices in New Jersey. Bulkley has reviewed articles for *Educational Evaluation and Policy Analysis*, *Educational Policy*, and *Policy Studies Journal*.

Copyright 2001 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Sherman Dorn
University of South Florida

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
hmwkhelp@scott.net

Thomas F. Green
Syracuse University

Alison I. Griffith
York University

Arlen Gullickson
Western Michigan University

Ernest R. House
University of Colorado

Aimee Howley
Ohio University

Craig B. Howley
Appalachia Educational Laboratory

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Benjamin Levin
University of Manitoba

Thomas Mauhs-Pugh
Green Mountain College

Dewayne Matthews
Western Interstate Commission for Higher Education

William McInerney
Purdue University

Mary McKeown-Moak
MGT of America (Austin, TX)

Les McLean
University of Toronto

Susan Bobbitt Nolen
University of Washington

Anne L. Pemberton
apembert@pen.k12.va.us

Hugh G. Petrie
SUNY Buffalo

Richard C. Richardson
New York University

Anthony G. Rud Jr.
Purdue University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Jay D. Scribner
University of Texas at Austin

Michael Scriven
scriven@aol.com

Robert E. Stake
University of Illinois—UC

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

other vols. | abstracts | editors | board | submit | comment | subscribe | search

This article has been retrieved 819 times since March 19, 2001

other vols. | abstracts | editors | board | submit | comment | subscribe | search

Education Policy Analysis Archives

Volume 9 Number 8 **March 19, 2001** **ISSN 1068-2341**

A peer-reviewed scholarly journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2001, the **EDUCATION POLICY ANALYSIS ARCHIVES**.
Permission is hereby granted to copy any article
if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in
the *Current Index to Journals in Education*
by the ERIC Clearinghouse on Assessment
and Evaluation and are permanently archived
in *Resources in Education*.

The Effects of Vouchers on School Improvement: Another Look at the Florida Data

Haggai Kupermintz
University of Colorado at Boulder

Related article.

Abstract
This report re-analyzes test score data from Florida public schools. In response to a recent report from the Manhattan Institute, it offers a different perspective and an alternative explanation for the pattern of test score improvements among low scoring schools in Florida.

Introduction

A recent report from the Manhattan Institute think tank (Greene, 2001) examined test scores of Florida public schools in 1999 and 2000 to determine the effects of vouchers on student performance. The report ends with a conclusion: "The most plausible interpretation of the evidence is that the Florida A-Plus system relies upon a valid system of testing and produces the desired incentives to failing schools to improve their performance." My own analyses of the Florida data lead to no such conclusion. Instead, I found the evidence telling a more interesting, and to my mind a more believable, story. I will argue that the evidence suggests that the "voucher effect" follows different patterns in the three tested subject areas: reading, math, and writing. Moreover, I will show that the most dramatic improvements in failing schools were realized by targeting and achieving a minimum "passing" score on the writing test, thereby escaping the threat of losing their students to vouchers.

Background

The Florida A-Plus school accountability program is based on tracking schools' performance and progress toward the educational goals set in the Sunshine State Standards. The main source of information on school performance is a series of standardized test in reading, math, and writing, known collectively by the somewhat redundant name FCAT (Florida Comprehensive Assessment Tests). All elementary, middle, and high school students are tested annually (different subjects in different grades) and the results are used to assign a grade to each school, from A to F, according to a formula that weighs the number of students performing below and above pre-defined markers along the test score scales. An F grade assignment has a variety of consequences and a great deal of attention is directed toward F schools in the Florida system.

One of the most visible and politically contested consequences of failing the State's tests is the voucher provision. If a school received another F grade in a four-year period, its students become eligible to take their public funding elsewhere to a private or better-performing public school. In 1999, 78 schools have received an F grade. Greene's report examines the gains these schools made on the FCAT between 1999 and 2000, and the executive summary offers a précis of the evidence: "The results show that schools receiving a failing grade… achieved test score gains more than twice as large as those achieved by other schools. While schools with lower previous test scores across all state-assigned grades improved their test scores, schools with failing grades that faced the prospects of vouchers exhibited especially large gains" (Greene, 2001, p. ii). The report itself compares the average score gains of higher-scoring F schools to lower-scoring D schools, serving as a control group. Standardized group differences constitute Greene's estimated effect sizes of the "voucher effect"—0.12 in reading, 0.30 in math, and 0.41 in writing. Other analyses in the report calculate the correlations between FCAT and other standardized test administered in Florida schools, to gauge the validity of the FCAT.

These findings lead Greene not only to the conclusions cited above, but also to strong public commentary in the local and national press in favor of Florida's voucher system and similar proposals in President Bush's school reform plan. The moderate "voucher effect" estimates and relatively cautious language of the report were replaced in the media by strong statements, emphasizing the magnitude of the raw score gains achieved by F schools. In an interview to the St. Petersburg Times (February 16, 2001), after the release of his report, Greene asserted: "The F schools showed tremendous gains because they faced a particularly concrete outcome that they wished to avoid: embarrassment, loss of revenue, vouchers". Even more boldly, generalizing from the Florida findings, Greene offered the following proclamation in a guest commentary in The New York Post (February 21, 2001): "So the improvement by Florida's failing schools was real. So, as debate proceeds over President Bush's education proposals, know this: Testing, accountability and choice are powerful tools to improve education - and, in particular, to turn around chronically failing schools. That's not a theory, but proven fact."

My re-analyses of the Florida data suggest that Greene might have over-stated the case for the simple explanation he promoted in his report and in the press. A more careful examination of the patterns of gains reveals that failing schools responded with a more sophisticated strategy than the undifferentiated, gross "voucher effect" gave them credit for. The key element of the strategy was to achieve a particular score on the writing test, in order to elevate their grades. The strategy was extremely successful and all failing schools were able to escape the threat of vouchers by achieving a grade of D or better in 2000.

Data

The data for the analyses are school mean scores on the FCAT reading, math, and writing tests from 1999 and 2000. They include all curriculum groups in both years (available on-line from the Florida Department of Education web site: http://www.firn.edu/doe/sas/fcat.htm). These data are slightly different from the data Greene used in his analyses, but as he comments (Greene, 2001, Note 10), the difference is inconsequential and similar conclusions will be reached using either dataset. The analyses below address issues that Greene either paid no attention to in his report or dismissed as unimportant. The first example of the latter is regression toward the mean.

An elusive regression artifact

On page 10 of his report, Greene alerts his readers to the potential biasing affect of regression to the mean:

> As another alternative explanation critics might suggest that F schools experienced larger improvements in FCAT scores because of a phenomenon known as regression to the mean. There may be a statistical tendency of very high and very low-scoring schools to report future scores that return to being closer to the average for the whole population. This tendency is created by non-random error in the test scores, which can be especially problematic when scores are "bumping" against the top or bottom of the scale for measuring results. If a school has a score of 2 on a scale from 0 to 100, it is hard for students to do worse by chance but easier for them to do better by chance. Low-scoring schools that are near the bottom of the scale are very likely to improve, even if it is only a statistical fluke.

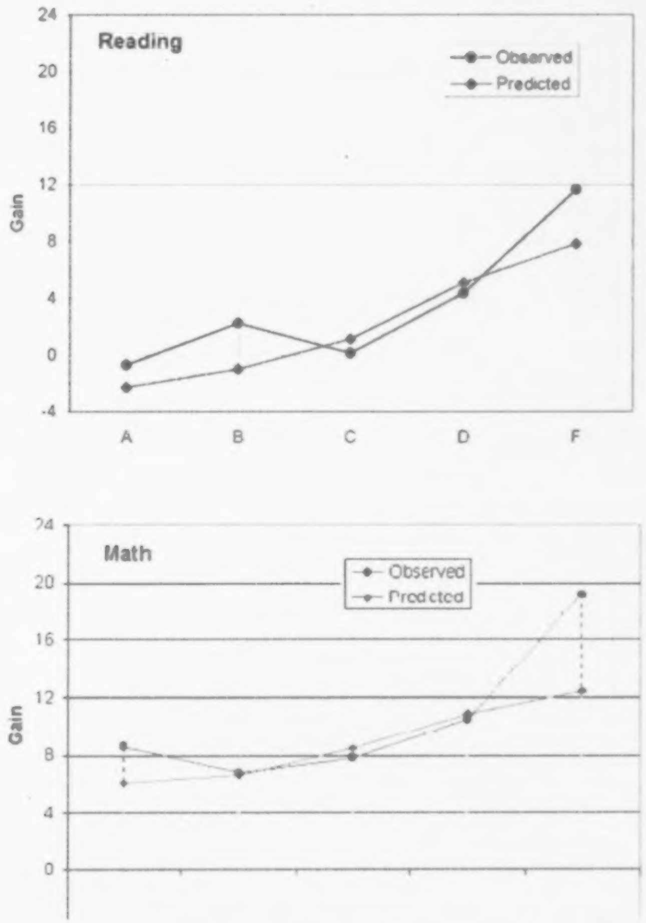
He then dismisses the threat because "the scores of those [F] schools were nowhere near the bottoms of the scale of possible scores" (p. 10). Greene seems to confuse regression toward the mean with floor and ceiling effects–completely different phenomena. Scores "'bumping' against the top or bottom of the scale" colorfully characterizes ceiling and floor effects but is an inadequate description of the regression effect. Regression toward the mean operates whenever the correlation between two variables (the 1999 and 2000 test scores, in our case) is less than perfect. It influences the entire range of scores—not just the very extreme—with a force proportional to their distance from the sample mean. Therefore, the fact that F schools where far from the bottom of the score scale is a poor indication that regression effects are absent. The two relevant pieces of information are how far the group is *from the sample mean* and the magnitude of the correlation between the two variables involved. Knowing these two quantities allows us to forecast the expected magnitude of the pull toward the sample mean. Using standardize scores aids interpretation, as the predicted standardized Y equals $Zy = rZx$ (X and Y are the 1999 and 2000 test scores, respectively). For example, a school 2 standard deviation below the mean in 1999 will be expected to score only .85(2) = 1.7 standard deviations below the mean in 2000, assuming a correlation of .85 (a value compatible with the typical correlation is the Florida data)—an effect size of .3! In 1999, F schools were 1.9*SD*s below the mean in reading, 1.7*SD*s below the mean in math, and 1.8*SD*s below the mean in writing. This simple analysis shows that the excepted magnitude of the regression effect warrants serious attention.

Using a slightly more complicated formula (see, e.g., Campbell & Kenny, 1999, p. 28, Table 2.1), and the regression coefficient instead of the correlation, one can calculate the expected 2000 score or the expected score gain, given a particular level of performance in 1999. Table 1 gives the expected score gains, if regression toward the mean was the only factor responsible for these gains, for the three FCAT tests, alongside with the observed gains for schools with different grades in 1999 [Note 1]. Figure 1 shows the same findings graphically.

Table 1
Predicted and Observed Gains By School Grade

	Reading		Math		Writing	
Grade	Observed	Predicted	Observed	Predicted	Observed	Predicted
A	-.68	-2.29	8.62	6.11	.24	.27
B	2.24	-1.01	6.85	6.65	.27	.29
C	.15	1.13	7.83	8.47	.29	.30
D	4.37	5.12	10.47	10.90	.33	.33
F	11.64	7.81	19.18	12.42	.67	.37

Figure 1. Predicted and Observed Gains By School Grade



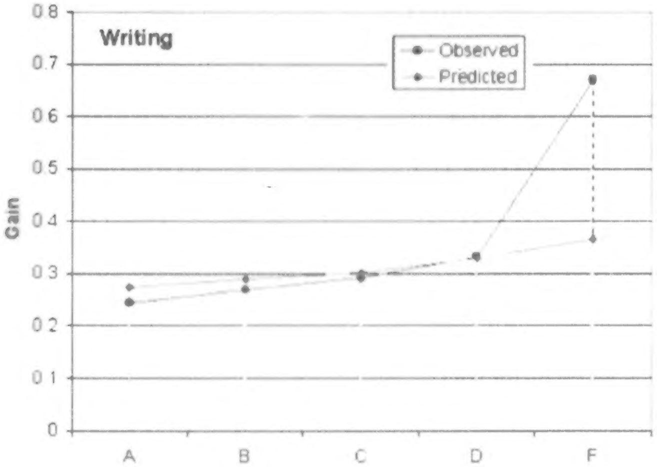


Figure 1 portrays an interesting picture. The height of each red dot represents the observed gain in scores between the 1999 and 2000 administrations of the FCAT. The blue dots represent the predicted gains attributed to the regression effect, and the distance between the red and blue dots, connected by a dashed line, depicts the "residual gain"—the amount of gain left after the regression effect has been accounted for. From Figure 1 we learn that a substantial portion (67% in reading, 64% in math, and 55% in writing [Note 2]) of the observed gains among F schools is due to regression to the mean. Note also that F schools do not appear exceptional and their residual gains are comparable to those observed in B schools, for example. These schools, however, start to stand out when we examine the patterns in math and even more so in writing. These observations agree with the order of effect sizes reported by Greene in Table 3 of his report. Unfortunately, Greene stops here to conclude: "a voucher effect." But the story has just begun to unfold.

Within-group patterns

We now direct our attention to the patterns of change within each group of schools designated by the same grade. In his second response to the potential regression threat, Greene suggested that "if the improvements made by f schools were concentrated among those F schools with the lowest previous scores, then we might worry that the improvements were more of an indication of regression to the mean (or bouncing against the bottom) than an indication of the desire to avoid having vouchers offered in failing schools". Curiously, while Greene argues for this strategy he never conducts the analysis. Instead, he presents in Table 5 *residual gains* that already take the regression effect into account. Even then he ignores the large difference between lower and higher scoring F schools in writing. Ironically, this difference is 0.16, exactly equal to the "voucher effect" in writing! Moreover, the same rationale for using residual gains here should apply with equal force for the gains reported elsewhere in Greene's report. The basic logic remains the same between tables.

Figure 2. Observed Gains by Initial Status and School Grade

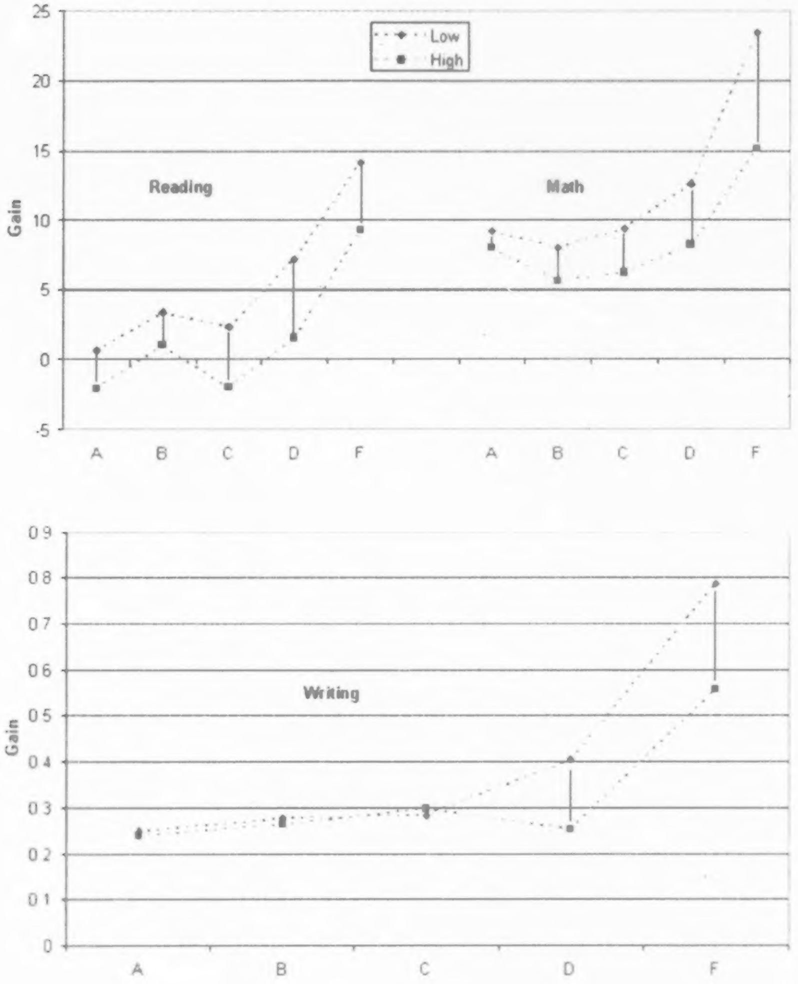


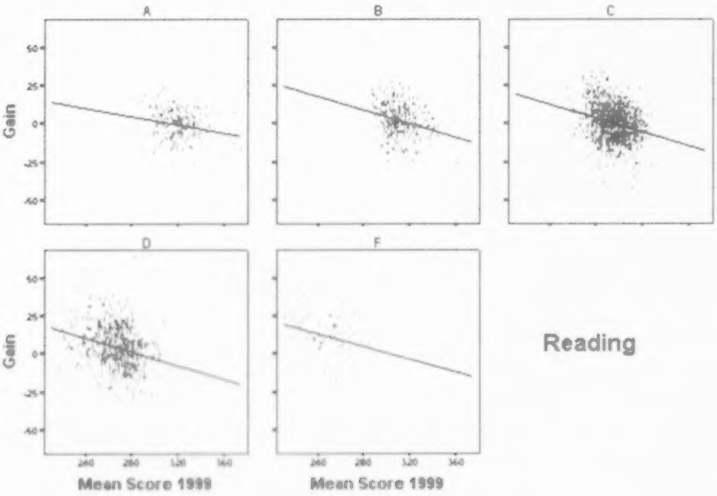
Figure 2 might cause us to worry, as Greene was right to point out. The red dots are the average gains made by the lower scoring schools (below the group median [Note 3]) and the blue dots the average gains made by higher scoring schools (above the group median) in each grade group. While the differences between gains of lower and higher scoring schools are constant across grade groups for reading, they increase substantially as grades get lower for math. For writing only, D and F schools show within-group differences, and these are more pronounced among F schools. In fact, the difference between higher and lower scoring F schools in writing is 0.23 representing an effect size of 0.23/0.39 = 0.6, substantially larger than the largest voucher effect Greene reports (an effect size of 0.41 in writing, see Table 3 in Greene's report)!

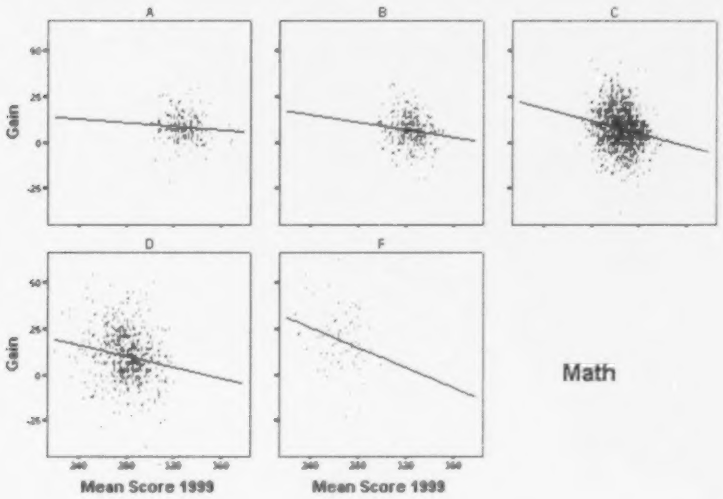
The within group analysis needs to be refined further as we change lens to zoom in on the details of patterns of gains within the different grade groups. Figure 3 shows the scatter plots of the 1999 and 2000 scores with the linear fits superimposed and depicting the overall trends in the data. Table 2 complements the graphs by giving the standardized regression coefficients corresponding to the trend lines.

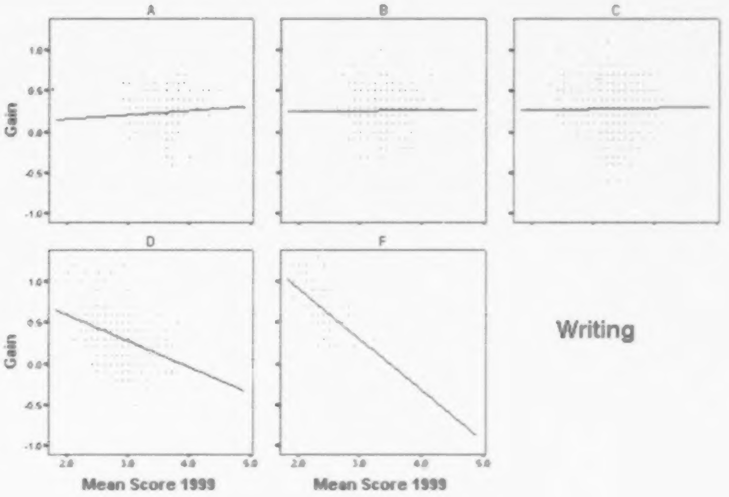
Table 2
Standardized Regression Coefficients of Gains Predicted from 1999 Scores

Grade	Reading	Math	Writing
A	-0.23	-0.09	0.07
B	-0.26	-0.14	0.01
C	-0.27	-0.20	0.02
D	-0.28	-0.19	-0.39
F	-0.28	-0.26	-0.54

Figure 3. Gains as a Function of 1999 Scores by School Grade





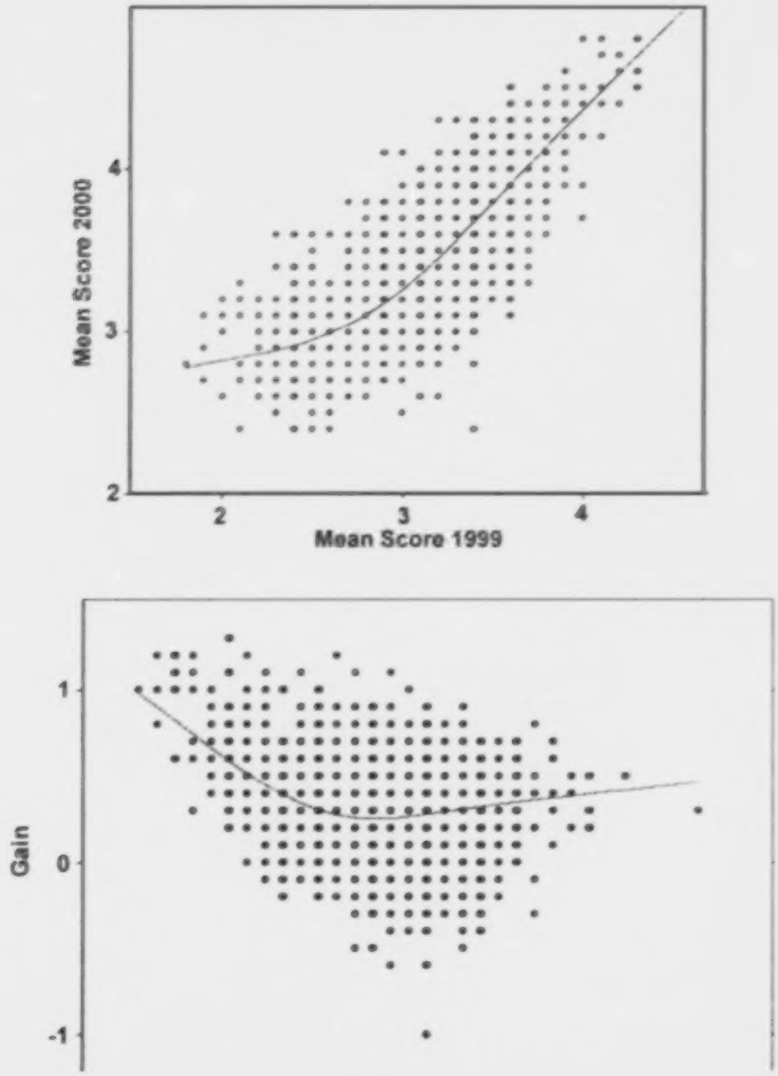


The reading scores behave as expected—a moderate negative correlation in all grade groups between the score achieved in 1999 and the gain realized one year later. Consistent with the patterns we identified in the cruder comparisons of Figure 2, the link between prior scores and gains becomes stronger as grades go down, a pattern most pronounced in writing. The findings for writing are striking. The amount of gain in F schools, and to a lesser extent D schools, is strongly determined by how low their scores were in 1999; the standardized regression coefficient is -0.54, representing the effect size of the mean gain difference for schools that scored one standard deviation apart from each other in 1999 (closely resembling the effect size value for lower and higher scoring F schools we calculated before). This pattern is completely absent for A, B, and C schools, whose 1999 scores provide no information on their expected gain.

The writing on the wall

The seemingly curious pattern of gains for writing has, in fact, a simple explanation. If there was a clear mark on the writing score scale that D and F schools set up to reach, not more nor less, then lower scoring schools would have to close a wider gap to reach the mark, giving rise to a strong negative correlation between where they started and how far they had to go (their gain). Figure 4 clearly demonstrates this phenomenon. It shows, for the entire school population, the relationships between 1999 scores and 2000 mean scores and gains. The lines represent the best fitted nonlinear trend lines (using the "loess" technique, see Chambers & Hastie, 1991, pp. 309-376).

Figure 4. Writing 2000 Scores and Gains as a Function of 1999 Scores



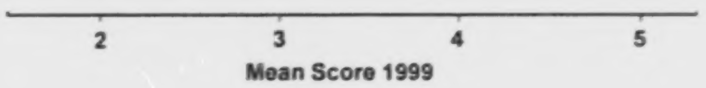


Figure 4 strongly suggests that the mark was a score of 3.0 on the writing test. Schools who scored less than 3.0 in the 1999 assessment have managed to make up the difference and reach the mark in 2000. The gain slope starts an upward bend below 3.0 in 1999—schools that scored less than 3.0 in 1999 have stabilized their performance around a score of 3.0 in 2000.

Conclusion

On June 21, 2000, long before the release of the Manhattan Institute report, the St. Petersburg Times ran a story entitled "Why are Florida children writing so much better?" Noting the impressive improvement in the writing score, the story offered an explanation: "How could so many kids suddenly become competent writers? Many educators were not completely surprised at the improvement. Out of fear and necessity, Florida educators have figured out how the state's writing test works and are gearing instruction toward it—with constant writing and, in many cases, a shamelessly formulaic approach. For some struggling schools, the writing test has helped them avoid an F rating." My findings are consisted with this explanation.

The pattern of score improvements on the FCAT ought to give Florida officials pause and trigger a serious research effort to identify potentially harmful imbalances and deficiencies in the A-Plus program. Until a far better understanding of and experience with the Florida accountability system is at hand, Greene's brave generalization from the Florida data he examined to the desirability of a nation-wide implementation is premature at best. It appears that the program's strong attention to the lower portion of the score distribution and the aggressive efforts to improve test scores in that region have produced substantial unintended consequences. Much more evidence is needed to arrive at a sufficiently detailed account of the program's operations and impact. The short list will include documentation of instructional practices in response to the incentive system in place for high and low scoring schools; an examination of the implementation and utility of school improvement plans; and data on possible program effects on retention, drop-out, and inter-school mobility patterns.

If vouchers were a dominant influence in motivating failing schools to act, the action they produced cannot be considered desirable by anyone who aims to "raise the bar" for students and schools. A minimum performance level in writing should not be considered a worthy educational goal for an ambitious accountability system such as the Florida A-Plus program. Yet, this appears to be the main achievement of the program in F schools. Coupled with a pattern of stagnation in other grade groups, especially in reading, these findings point to aspects of the program that deserve closer scrutiny. However, the reader of the Manhattan Institute laudatory report is offered a false sense of a dramatic success. It is, therefore, appropriate to recall Cronbach's advice to the evaluator:

> Disillusion is the bitter aftertaste of saccharine illusion. It is self-defeating to aspire to deliver an evaluative conclusion as precise and as safely beyond dispute as an operational language from the laboratory… . When the evaluator aspires only to provide clarification that would not otherwise be available, he has chosen a task he can manage and one that have social benefits. (Cronbach, 1980, p. 318)

Notes

1. The calculations of the regression coefficients in these analyses excluded F schools to avoid attributing a potential true program effect to the regression artifact.
2. These percentages are calculated as the observed gain divided by the predicted gain and multiplied by a hundred. For example the figure for reading is (7.81/11.64) x 100 = 67%.
3. The choice between the mean and median is inconsequential in this analysis. I used the median because it produces slightly more equal sample sizes.

Acknowledgment

The work reported here was supported under the Educational Research and Development Centers Program, PR Award Number R305B600002, as administered by the Office of Educational Research and Improvement, U.S. Department of Education. The finding and opinions expressed in this report do not reflect positions or polices of the National Institute on Student Achievement, Curriculum, and Assessment, the Office of Educational Research and Improvement, or the U.S. Department of Education. My thanks go to Greg Camilli, Sherman Dorn, Steve Lang, Bob Linn, Lorrie Shepard, and Kevin Welner for helpful comments.

References

Campbell, D. T., & Kenny, D. A. (1999). *A primer of regression artifacts*. New York: Guilford Press.

Chambers, J. M. and T. J. Hastie, Eds. (1991). *Statistical models in S.* Pacific Grove, CA: Wadsworth & Brooks/Cole.

Cronbach, L.J. and Associates. (1980). *Toward reform of program evaluation.* San Francisco CA: Jossey-Bass

Greene, J. P. (2001). *An Evaluation of the Florida A-Plus Accountability and School Choice Program*. New York: The Manhattan Institute.

About the Author

Haggai Kupermintz
School of Education
University of Colorado at Boulder

Email: haggai.kupermintz@colorado.edu

Haggai Kupermintz is an Assistant Professor of research and evaluation methodology at the University Colorado at Boulder, School of Education. His specializations are educational measurement, statistics, and research methodology. His current work examines the structure, implementation, and effects of large-scale educational accountability systems.

Copyright 2001 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple **University of Wisconsin**	Greg Camilli **Rutgers University**
John Covaleskie **Northern Michigan University**	Alan Davis **University of Colorado, Denver**
Sherman Dorn **University of South Florida**	Mark E. Fetler **California Commission on Teacher Credentialing**
Richard Garlikov **hmwkhelp@scott.net**	Thomas F. Green **Syracuse University**
Alison I. Griffith **York University**	Arlen Gullickson **Western Michigan University**
Ernest R. House **University of Colorado**	Aimee Howley **Ohio University**
Craig B. Howley **Appalachia Educational Laboratory**	William Hunter **University of Calgary**
Daniel Kallós	Benjamin Levin

Umeå University
University of Manitoba

Thomas Mauhs-Pugh
Green Mountain College

Dewayne Matthews
Western Interstate Commission for Higher Education

William McInerney
Purdue University

Mary McKeown-Moak
MGT of America (Austin, TX)

Les McLean
University of Toronto

Susan Bobbitt Nolen
University of Washington

Anne L. Pemberton
apembert@pen.k12.va.us

Hugh G. Petrie
SUNY Buffalo

Richard C. Richardson
New York University

Anthony G. Rud Jr.
Purdue University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Jay D. Scribner
University of Texas at Austin

Michael Scriven
scriven@aol.com

Robert E. Stake
University of Illinois—UC

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García

Angel Ignacio Pérez Gómez

(México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky
(Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

(Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres
(U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

other vols. | abstracts | editors | board | submit | comment | subscribe | search

This article has been retrieved 524 times since March 21, 2001

other vols. | abstracts | editors | board | submit | comment | subscribe | search

Education Policy Analysis Archives

Volume 9 Number 9 **March 21, 2001** **ISSN 1068-2341**

A peer-reviewed scholarly journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2001, the **EDUCATION POLICY ANALYSIS ARCHIVES**.
Permission is hereby granted to copy any article if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

The Academic Journal:
Has it a Future?

Gaby Weiner
Umeå University, Sweden

Abstract
This article examines the current state of the academic journal. It does so for a number of reasons: the increasing expense of paper journals; the advent of electronic publishing; the use of publication in journals as an indicator of research quality (in addition to disseminating knowledge within a discipline) and consequent criticisms of systems of peer review and evaluation of scholarship; emergent issues of equity and access; and evidence of malpractice. These issues taken together constitute a critique of, and challenge to, the process whereby research papers become journal articles, which has in the past been viewed as unproblematic and straightforward. This paper brings together a wide range of literature in order to

inform discussion about the future of the academic journal. It briefly examines the origins of the academic journal and then provides a comprehensive overview of current debates concerning how academic journals work today. In so doing, it raises questions about decisions that will need to be taken regarding the continuity or otherwise of the conventional academic journal, and how publishing practices may change in the future.

This journal, *Education Policy Archives Analysis*, available online, free of charge, and produced with minimal maintenance costs, is indicative of why scholarly publishing is in crisis. The future of paper journals has been put in doubt by the emergence of the electronic journal, of which there were 1,465 in 1997 (Association of Research Libraries, 1998). The "Communication of Research" Special Interest Group of the American Educational Research Association maintains a directory of freely accessible, peer-reviewed scholarly journals in education, of which there were 93 available as of February, 2001: http://aera-cr.ed.asu.edu/. Paper journals are also threatened by other forces, for example, the proliferation of paper and electronic journals as a result of the "publish or perish" academic cultures of many western countries, and the increased use of the academic journal as a means of evaluating the quality of one's scholarship. The widespread introduction of research reviews and assessment exercises based largely on publication in learned journals has led to perceptions that the practices of academic journals are more important to individual academics and their institutions than ever before. Thus, criticisms have been raised regarding the use of published work as an academic "performance indicator" and about the need for standard, equitable and open journal procedures and practices. Assurance has been asked, for example, that papers are dealt with fairly and that different journals use similar procedures and criteria for submitted manuscripts.

Eisenstein (1979) tells us that two potentially incompatible processes of change ushered in the first print revolution in the 1450s: one "gradual and evolutionary" and the other, "abrupt and revolutionary":

> Thus the invention and utilization of movable type may be viewed as one by-product of previous developments, such as the spread of lay literacy, and as a factor, which, in turn, helped to pave the way for later developments, such as modern mass literacy. (p. 33)

A similarly significant challenge to movable print is now with us, this time from electronics and telecommunications. This brings with it clear signals that the dominance of the paper journal, the main form of academic knowledge communication for the five centuries since

Gutenberg, may be coming to an end. Whether the conventional form of paper academic journal is viable, necessary, effective or affordable in the present economic context is in some doubt.

Yet, even though some academics (and librarians) have become critical of today's system of academic publishing, others show few signs of dissatisfaction, and, indeed, seem ever more interested in strengthening their ties with publishers, both as producers and consumers. As a recent review of the state of academic publishing notes:

> What gives this enterprise its peculiar cast is the fact that the producers of knowledge are also its primary consumers. In most fields the market for scholarly publications is driven largely by the internal mechanics of a culture, in which further specialisation increases greatly the volume of published work at the same time as individuals come to read more narrowly within their field. (PHER, 1998:3).

Here I seek to clarify some of these issues by providing an overview of debates and studies concerning the role and impact of the academic journal. First, I explore the origins of the academic journal and how early traditions continue to influence academic journals today. Then I will attempt to map the range of debates in recent years among researchers and writers interested in academic publishing and its changing role. This article ends with a discussion on the future of the academic journal, and what changes are needed if it is to continue to be the main vehicle for academic communication.

The impact of Gutenberg was not immediately evident and in fact printers and scribes continued to copy texts manually for more than fifty years after the first moveable-type printing press was established: "one must wait until a full century after Gutenberg," Eisenstein notes,"before the outlines of the new world pictures began to emerge into view." (Eisenstein, 1979, p. 33) Writing in the middle, as it were, of another kind of revolution, this paper explores the various pulls for and against change in the context of academic publishing, but of course, can only but speculate about the eventual and extent of the outcomes.

The Origins of the Academic Journal: two traditions

There is some disagreement about the origins of the academic article depending on discipline. Reports. The first two scientific journals appeared in 1655: *Journal des sçavans* in France, and *The Philosophical Transactions of the Royal Society,* in England (Swales, 1990; Vrasides, 2000). The genre of the scientific article followed on from letters that scientists wrote to each other and thus many of the earliest contributions used the first person, as in the case of letter-writing. The aim of *Transactions* and other similar publications was to provide a general forum for discussion which eventually became transformed into a new genre of scholarly writing.

An additional, and powerful influence came from the convention of publishing scientific treatises in order to establish a sound foundation for scientific knowledge. To establish the factual nature of experimentation, mid-seventeenth century scientists such as Robert Boyle, developed "a largely self-conscious and highly complex set of strategies" (Swales, 1990, p. 111). This involved making public the form of the apparatus used (actual or by detailed drawings) and if possible carrying out the experiment in front of an audience—so that agreement of the relevant community could be gained. Replication of experiments was also believed to strengthen any scientific claims, though clearly experiments had to be successful to do so. Written accounts of experiments were lengthy and detailed so that readers could feel they were gaining a true account, whether or not the experiment succeeded. Claims were deliberately cautious and philosophical speculation was avoided. Bazerman's (1983) study of the development of the *Transactions* during the period 1665-1800, however, shows that the articles were neither uniform nor were they mainly experimental. In the early days of the journal, the majority of reports were of "natural" phenomena such as earthquakes, or anatomical observations and dissections. Later, understanding of the complex nature of phenomena led to a more uniform approach.

> In this process of evolution, the scientist's relationship with nature gradually changed from a view that the nature of things would be easily revealed by direct or manipulated observation to a view that nature was complex, obscure and difficult to get at. Inevitably enough, this changing view also meant that more care began to be taken in describing how experiments were done, in explaining why particular methods were chosen, and in detailing precisely what results were found (Swales, 1990, p. 113).

The humanities took a different later pathway to the scholarly article. Today's scholarly journals are modelled on those developed for the new "professional history" of nineteenth- century Germany (McDermott, 1994). One of the first historical periodicals, still in existence, is the *Historische Zeitschrift* which appeared in 1859, some two centuries after the first scientific journals (Steig, 1986). Based in universities which were regarded as central and unifying institutions of academic professionalism, scholarly journals in the humanities were used in Germany to bring coherence to a discipline, and as a means of communicating knowledge among like-minded scholars. Ideological commitment was considered congruent with scholarship; and political discussions were included alongside more recognisable academic contributions. The conviction that politics is incompatible with scholarship became widespread only after the Nazis took control of German universities in the 1930s: hence the post-war emphasis in

Historische Zeits[illegible]ft on "the maintenance of rigorous scholarly striving towards true unbiased knowledge" (Steig, 1986, pp. 134-5).

The legacy of the two traditions for today's academic writing remains evident, causing much debate among those who have sought to unify and generalise across disciplines. This has often confused students and beginning researchers who have questioned whether it is "more scholarly" to use the first or third person in academic writing or whether all research articles need to follow a standard "scientific" form. Or indeed whether it is so necessary to take up a stance of neutrality and objectivity. And, of course, there are as many responses as questions, all highly dependent on specific disciplinary and research cultures.

Today's Academic Journals

Despite academic publishing's distant and relatively modest origins as described above, it has enlarged and diversified, conventionally embracing a wide number of forms: for example, books of varying lengths written by one or more authors; collections of articles edited by one or more academics; research monographs or reports; undergraduate and postgraduate texts; vanity (i.e., self-financed) monographs or books; articles in regular or special issues of journals, and so on. The *academic journal*, however, is distinctive from other forms of publishing in certain key ways. It is likely to be university-based; it involves academics editors and consultants; it uses standard forms and styles of binding, type-setting and publishing; and it is published at regular intervals (McDermott, 1994). Furthermore, academic journals usually employ referees, that is, experts in specific fields, who are asked to comment and make recommendations as to whether submitted manuscripts merit publication.

Academic journals are used in three main ways: first and still most importantly, to produce, disseminate and exchange academic knowledge; second to rank research and scholarly work in order to aid the distribution of research funds; and third, to inform decisions concerning appointment and promotion. The second and third factors, in particular, have meant that journals and the procedures they use have become more important to individual writers and academics, and their institutions. This is most acute where research activity is highly prioritised and where it constitutes a significant source of institutional income.

However, to understand how academic journals work, it is also important to understand that they have at their core a set of social, economic and academic relationships which involve a complex variety of roles and people. At different times, individuals may hold positions and responsibilities for different journals at the same time. They may, for example, be editor, editorial board member or referee for one or more journals at the same time as trying to get a paper published in these or other journals.

A useful way of looking at academic writing is as a social game, the rules of which need to be understood before individuals are able to successfully engage with it. For example, Clark and Ivanic use the term "literacy practices" to include both the social conventions and "the physical, mental and interpersonal practices that constitute and surround the act of writing" (Clark and Ivanic, 1997, p. 12). Hence we may refer to the "literacy practices" of academic journals (meaning both the practices employed in researching and writing papers, and the social rules and regulatory frameworks surrounding them) when we explore similarities or differences between academic journals within the same discipline or between disciplines. "Practices" are largely determined by dominant individuals or groups at any historical moment, although writers have the option, in principle, not to conform to given practices if they so wish. Thus power is important in writing since the need for acceptance shapes practices of both form and content. But power can also be used in another way —as in "the power of writing." The writing act itself is associated with great power —it can provide access to influence over others through the communication of ideas and the use of rhetoric, which, in the case of the great philosophers, playwrights and novelists, can endure for hundreds of years.

Another useful concept is "discourse community" which if applied to academic disciplines and sub-disciplines helps explain why, until now, there has been relatively little disagreement about how academic journals work. In order to enter and be part of a particular discourse community, individuals need to share certain characteristics. These include: a broadly conceived set of public goals; mechanisms for communication between members and circulation of information and feedback; utilisation of specific language practices; and membership requiring a level of specific expertise and knowledge-base. Such a concept of "discourse community" shows what binds specific groups of academics together, how others come to be excluded, the relative conservatism of such communities, and the potential difficulty of introducing changed practices (Swales, 1990). However such communities are also sites of contestation which may lead to break-away sub-disciplines generating new discourse communities (and new journals).

The power of certain groups ("experts") to shape and confirm the production of certain kinds of knowledge determines the ethos and membership of each discourse community. As a consequence, "outsider" or unofficial knowledge may be disqualified and dismissed as non-rigorous, undisciplined, and unprofessional. In his conceptualisation of power/knowledge configurations, Foucault (1980) focused on the power of research to control as well as to generate knowledge. This does not mean that oppositional viewpoints are eradicated: rather the inclusion of different (but tolerated) viewpoints not only confirms academics' espoused commitment to freedom of speech and respect for diversity of opinion, but indicates the boundaries and limitations of what may be said and written. Thus, as Apple states, "reproduction and contestation go hand in hand." (Apple, 1982, p. 8).

Challenges to Academic Journals

A number of developments have taken place in recent years that challenge the foundational paradigm of the conventional academic journal. Considered in this section are the economy of journals, the impact of electronic journals like *Education Policy Archives Analysis*; peer review and the assessment of research productivity and quality; and social justice and ethical issues.

The Economy of Journals

The conventional academic journal has been highly profitable for publishers, because copy is consistently produced (with copyright assigned to the publisher) while academics generally give their labour free—as writers, reviewers, editors and members of editorial boards. The paradox is that on the one hand, academic institutions make the initial outlay in the form of salaries and infrastructure to support the research which provides the raw material for articles and to provide editorial labour for the journals: on the other, universities, colleges and individual academics are made to pay heavily (through subscriptions) for the publication and distribution of that research.

The act of publishing has been referred to as "a gift exchange" within a community of like-minded people —where the gift, freely given, generates esteem and professional advancement (PHER, 1998:3). However the producers are not held responsible for market failure, neither are they beneficiaries of market success. Rather their role is to keep the system fuelled by submitting papers, by providing academic editorial services, and as purchasers.

In their original conception, journals *belonged* to those who wrote for them and read them, being in the main published by university presses. This remained the case until the post- war period when, in the US in particular, the university sector expanded with an accompanying rise in level of publications from the increased number of academics in the system. Commercial publishers entered the scene at this point and were welcomed as one way to diffuse the bottleneck of papers waiting to be published. However publishers were quick to exploit the opportunities presented to them.

Recognising the bottleneck, commercial publishers came to absorb an increasing share of the market, with broad support of higher education institutions, scholarly societies, and faculty who served as editors, reviewers, and members of editorial boards. Consigning the production and distribution functions to the commercial sector purchased an immediate increase in capacity: existing journals expanded, and new journals were formed to accommodate a growing quantity of research in increasingly specialised domains (PHER, 1998:3).

Initially these arrangements seemed to work well, providing benefits for all concerned. Academics were able to get their work published, publishers took responsibility for the organisation and distribution of the journals, and profit margins seemed acceptably balanced against the cost of the journals. However problems began to emerge as the requirements of the market clashed with the academic milieu. For example, publishers required authors to turn over their copyrights and were thus free to buy and sell academic knowledge as a commodity. The burgeoning costs of print and distribution were passed directly over to the purchasers of the journals, enabling publishing houses to accumulate substantial profits. Thus, the British entrepreneur Robert Maxwell made his fortune in the 1970s and 1980s through the journals associated with his publishing house Pergamon. Academics, conventionally unworldly about financial matters, were slow to realise what was happening and the pressure to publish meant that they were willing collaborators in a system which exploited them.

Thus it comes as no surprise that the volume and price of academic information dissemination increased nearly three-fold in a decade with the "cost of scholarly journals increased [by] a whopping 148% in the US between 1986 and 1996 (PHER, 1998, p. 1 - 2). Concerns were raised about whether the creation of more and more knowledge outlets (through the creation of new journals) is indeed a solution. Indeed, the proliferation of new journal titles attracted criticism in the UK, both about the quality of much of the output of academic research and writing, and the problems quantity presents to the academic reader (Hillage et al., 1998).

The system we have now was designed, and seemed to work best in, the academic world of the 1960s when academic and market interests coincided. It produced, for a time, a form of academic scholarly discourse in printed form serving higher education institutions and their staff in a fair and cost effective manner. However the fit seems less perfect in the much changed academic climate, four or more decades later. Increased necessity to publish in academic journals in an expanded university sector has generated further pressure, both to increase the number of journals available and on library budgets, in particular. Predictably perhaps, whilst both the numbers and prices of academic journals have increased as have individual subscriptions to journals, the number of articles that an individual academic reads on average each year has remained much the same (about 150 to 190 articles). Again, it is libraries that have most felt the burden of journal proliferation.

Publishers report that as the number of journals have increased, academics have not increased their personal subscriptions, but have instead relied upon the library, with most academics continuing to subscribe to between three and four journals. Publishers also report that scholars are purchasing fewer personal copies of scholarly monographs, which has helped contribute to smaller press runs and the current tenuous economic situation of the scholarly monograph (University of Austin, 1998, p. 1)

The system we have now is clearly at a crucial point—some might say in a state of collapse—with librarians in the forefront of calls for urgent change.

> Those librarians who help you decode Dewey's decimals are becoming unlikely warriors at the end of this decade. They have to. With large publishing conglomerates driving the prices of scholarly journals higher and higher, librarians find themselves spending more and more money to purchase fewer and fewer books. Their constituencies are concerned. Scanning the stacks, professors moan; brooding their budgets, the financial officers grumble. It's no wonder that many librarians are asking: Is there a better way? If you don't like the way journals are being published, why not do it your self? (Rambler, 1999, p. 1)

Librarians have had the fullest picture of a crisis-in-the-making; because of academics' greater reliance on libraries for the journals and books they cannot afford, because of libraries' diminishing resources and reduced budgets, and also because of their need to develop paper and electronic systems simultaneously.

Electronic Publishing

An important challenge to the conventional paper journal has come from electronic publishing, as has already been noted - that is, the "full-blown usage of networked computers" (Waaijers, 1997, p. 77). The so-called electronic revolution emerged because of two main technological changes:

> First the evolved computer, now cheap, robust and powerful, and second, our recent ability to store and send huge quantities of data from computer to computer hither and thither across the globe by connections such as the internet. (Young, 1996, p. 290)

As electronic journals pioneered new forms of text production designed to reach a wider and more diverse readership, conventional academic journals continued as before. But demands for change came not only from the imperatives of technology. Pressures to incorporate electronic journals in current systems of academic publishing, and even to substitute them for paper journals, arose from a number of sources. For example, certain problems in the production of the paper journal are perceived as resolvable by electronic versions: in particular, the slowness of the process, proliferation of journals and high costs to university and college libraries. Electronic publishing makes possible faster turnarounds of papers from submission to publication and its potential to lower the production and distribution costs —by 30% or more – could lead to cheaper journals for libraries and individuals, although initial capital costs may be higher (Burbules & Bruce, 1995). Electronic publishing, moreover, creates possibilities for flexibility in the writer-reader relationship; with enhanced opportunities for interactivity, multiple-modes of data presentation, publication in more than one language and fewer restrictions on word-length and format (Vrasidas, 2000). Moreover, Glass (1999) claimed that online education journals also widen readership, to include groups such as teachers, administrators, school board members, and those living in countries, all previously unlikely to have access to scholarly literature

A less positive projection is that the promise of quick turnarounds may encourage hasty and under-developed submissions, and that lack of access to fast changing technologies of text communication is likely to increase exclusivity rather than wider access. Also if, as Glass (1999) suggested, "a reader in the year 2000 browsing a scientific journal from the year 1910 will find the environs thoroughly familiar," arrangements of storage and information retrieval in the new electronic era cannot promise such familiarity.

The term *archiving* denotes not only the storage of materials but the systematic organisation and exhaustive provision of access to these materials. In the case of electronic publications one of the major problems to be addressed in access provision has been the wide variety of formats in use. This was illustrated by the statement "I can read a printed book published 300 years ago but it is impossible for me to read a Microsoft Word II document written in 1988." (ICSU, 1998, p. 2).

Vrasidas (2000) neatly summarizes the range of reasons given against the broader acceptance of electronic journals.

> Among the most prevalent ones are the politics of controlling scholarly communication, the economic benefits of publishers, copyright issues, bandwidth issues, access to the Internet, the lack of skills to write for the web, the technology phobia among scholars, the prestige for publishing an online article versus an article in paper, and resistance to changing the old traditions of scholarly publishing that legitimizes the academic disciplines (Vrasidas, 2000, p. 4)

Notwithstanding, the advent of electronic publication has stimulated an extensive debate about conventional forms of journal publishing and whether the paper journal is now the most effective means of disseminating research and scholarship. It has provided a challenge to how the dissemination of scientific knowledge through journals is structured, and simultaneously, to existing systems of peer review.

Peer Review

The employment of peer review lies at the center of academic journals' procedures and practices. Each journal relies on the input of a panel of academics, each of whom has made a significant scholarly contribution to a particular field, and who is therefore assumed to be able to pass judgement on the quality of papers of colleagues and scholars working in the same or related fields. Ostensibly fair and non-hierarchical—what could be more non-hierarchical than being judged by one's equals?—nevertheless, the system is fraught with tensions, particularly where challenges are made to the reviewer's own work or academic stance.

Peer review has been chosen as the most just and appropriate means of coming to a decision about the quality of research, despite the recognised fallibility of some peer review systems and the consequent need to constantly review and reconsider their practices (ABRC, 1990). However, it has also drawn criticism for being inherently conservative, and a means by which powerful academics in a field (or within a particular discourse community) retain their grip on who contributes and what knowledge is generated. Because peer reviewers (known also as referees) are generally recruited through informal professional contacts, the system has also been condemned as an "old boy" network which is unfair to outsiders and newcomers (Furnham, 1990).

Another challenge to peer review has come from evidence both of substantial disagreement between referees when evaluating manuscripts and of lack of objectivity. This suggests, according to Berardo (1989):

> a differential application of established criteria and reflecting the biases of individual reviewers. There is little doubt that a reviewer's proclivities toward certain theoretical perspectives, methods of data collection and analysis, or substantive foci play a role in the evaluation process (Berardo, 1989, p. 133).

If evidence is available to support the view that the peer review process differs *within* a field or discipline as above, there is also evidence that differences can be found *between* disciplines. Harnad & Hamus (1997) suggest, for example, that variation in rejection rates does not necessarily indicate variations in scholarship.

In some disciplines, the mark of excellence is their rejection rate, which can be as high as 90% (and probably higher in a journal like *Science*); in other disciplines, it is the acceptance rate that is 90% or more—and this need not mean that the journal is of lower quality. Sometimes it is the very prestige of the journal that keeps contributors from submitting anything but their very best work to it for refereeing (Harnad & Hamus, 1997, p. 19).

Thus, we can see that while peer review is widely used by journals, it is more problematic than its widespread use suggests. As a system of accepting and rejecting papers within a discipline, peer review seems a reasonably robust strategy. However when the selection of papers is invested with different purposes, the discourse changes and becomes more complex - as we shall see with regard to the use of journals to evaluate research quality and productivity.

Productivity and Citation

Numerous and diverse methods have been developed to assess the quality of scholarship and rate of productivity of academics. However these are frequently complex and superficial as Hanish et al. suggest below.

> Productivity refers to the quantity of publications attributable to a given scholar, expressed in a lifetime total or a yearly rate when divided by the scholar's professional age. Impact generally means how frequently that an individual's work is cited by other authors, which likewise can be expressed as a lifetime total or a yearly rate. Quality is almost never assessed directly; productivity and impact, though, frequently pose in its place (Hanish et al., 1998, p. 1)

One of the most direct and straightforward measures of quality of work and research productivity is "the simple publication count" that is the number of publications an individual scholar has accumulated over a given period (Colman et al., 1992, p. 98). However, in the competitive climate of academia at the turn of the twenty-first century, merely to succeed in getting into print is not considered a sufficient guarantee of scholarship. Sometimes all publications are weighted equally. But how are co-authors to be accredited? Some assume an equivalent contribution from each author listed while others employ a weighting system based on authorship order (Hanish et al., 1998). There is also the issue of how to compare single-authored and co-authored work. Moreover, some journals "count" for more, for example, those included in citation indexes.

This brings us to an alternative method of evaluating scholarship—to count not publications but citations. The use of citation is premised on the assumption that the quality of a scholarly article can be gauged by the number of times it is cited in subsequent journal articles, books etc. Thus, a commonly used method of judging whether a particular academic journal or an individual scholar has made a significant impact on a field is to see how many times they have been cited by other scholars in the field. This has developed into a complex technology of measurement delivering "citation data" as "quantitative indicators" (Garfield, 1990) which can be used to evaluate existing journals and individuals against other journals and individuals, on a yearly or other chronological basis, and according to impact factor, i.e., whether citation occurred in a newspaper, article, research review and so on.

It is assumed that the higher the number of citations of an academic's work, the greater the peer esteem and therefore the higher the quality of scholarship (e.g., Field et al., 1991). In practice, the use of citations involves counting the number of citations over a specific period in journals covered by one or more of the established citation indexes —which raises a number of further problems. First, a large number of journals including the newest and most innovative, are absent from standard citation indexes. As Garfield (1990, p. 6) points out "no matter how many journals are on the market, only a small proportion account for most of the articles that are published and cited in any given year." Second, citation indexes are generally unable to distinguish between positive and approving citations, critical and dismissive citations, and self-citations. Third, citations too may be seen as merely reflecting the status quo, because of the frequency of self-citation and citation of friends (Field et al., 1991).

Whatever performance or quality indicator is used regarding publication, whether publication count or citation, a key factor for each institution in the present competitive climate is how the performance of its researchers measures up to others. Institutions which are able to prioritize investment in the buying in of productive researchers or in creation of a research milieu, are those most likely to see a positive outcome in terms of commercial or charity grants, or government funding. Put another way, there is a strong relationship between investment in research and its "quality" outcomes.

The most obvious output measures relevant to departmental research performance are simple publication counts and more elaborate publication-based measures designed to take quality into account. The most important input variables are the number of departmental staff members, the number of research assistants, the size of equipment and recurrent grants, and the amount of research income (Colman et al., 1992, p. 97).

When these performance indicators, however arrived at, are used as surrogates for the distribution of "quality" and "excellence," a crisis emerges not about selection but about social justice.

Equity and Access

At the annual meeting of the American Educational Research Association (AERA) in New York in 1996, the AERA Publications Committee noted that some inequalities relating to getting published lay outside its control and that perfect representation of authorship and content was impossible to achieve, despite strategies to increase diversity of authorship. In particular, the "struggle over hiring" in the US (such that proportionally few female or minority ethnic academic staff are appointed) has created preconditions which militate against greater inclusiveness in journals. The response of mainly young, graduate students on this occasion, however, was to be highly critical of existing publishing practices, in particular, what were seen as the lack of openness in the appointment of journal editors, lack of encouragement to new authors, and predominance of white/male networks of power.

AERA's response to these, and other similar points raised by its membership, was the development of a "list" of minority scholars, produced each year "for the purpose of increasing the availability, visibility, and representation of minority scholars within AERA's visible structure" to AERA division and committee chairs, journal editors etc. (AERA, undated). This has encouraged those in the most senior echelons of the US educational academic community to widen their conventional notions of whom to appoint to what —though it is difficult at the present time to estimate, with what success.

Thus we can see that the discourses of excellence, competitiveness and, to some extent, exclusivity which have suffused academic journals since their inception, have not necessarily provided a fruitful ground for discussion of social justice or equity issues. The exclusive nature of academia, indeed, is seen to underscore its claim to excellence. However, following developments of equity policies in other areas of academia (Weiner, 1998), who writes in academic journals has become a topic of considerable importance. Questions arise as to whether there is evidence of sexism, racism or other unjust practices in academic publishing and whether new forms of publication are likely to promote a change in publishing's ethos of elitism. Does electronic publishing favour the favoured, or does it enhance equality of access and usage?

Sociologists of science have suggested that certain characteristics of writers, for example, where they were educated and are presently employed, influence reviewers' recommendations and editors' decisions about whether or not to publish (Bakanic et al., 1987). Thus a "big" name may well gain the advantage in the competition for journal space in various ways:

> Judgement ....may be systematically skewed by deference, by less careful appraisals involving exacting criteria, by self-doubts of one's own sufficient competence to criticise a great [scholar] or by fear of affronting influential persons in the field (Zuckermann and Merton, 1971, p. 82)

Following feminist activity in other areas of academia, gender has recently received attention as a factor in academic scholarship and writing. An aim of feminist research into higher education generally has been to "generate a transformation of the academy" by highlighting discrimination and by developing theories and frameworks for gender difference (Townsend, 1993, p. 22). Gender studies of academic publishing have reported a number of consistent findings: that women or feminist issues rarely form the topic of mainstream journals, though there has been a slight increase in recent years (Townsend, 1993); that male authors have generally higher profiles and higher productivity than women, are cited more and are more likely to self-cite (Helmreich et al., 1980); and that male authors are more likely to be cited by men (Ward et al., 1992). However, Over showed nearly two decades ago that article-for-article, women are as likely as men to be cited, but their proportion of citations is lower because of their lower overall publication levels (Over, 1982). It should also be noted that there is a small, specialist group of publications focusing primarily on gender or women's issues, which draws a mainly female authorship and readership.

Other social patterns of authorship, for example, ethnic origin or colour, have attracted less attention although there is some evidence that minority and black writers are as under- represented as authors, as they are as a focus of study. In the latter instance, a study of the proportion of articles on minorities in psychology and education journals in the US between 1952 and 1973, found less than 2% discussing minority issues (Van Scoy & Oakland, 1991). It is likely to be minority and black researchers and academics who are most interested in exploring "minority" issues in research, if trends are similar to those of women researching and writing about gender issues. This suggests that there are relatively few minority and black academics as researchers and authors, although there may have been a slight improvement in numbers more recently.

Countering Malpractice

Another less visible issue for academic journals but one that has come to prominence for several different reasons in recent years, is ethical considerations regarding journals and intellectual property rights. It has been argued that the intense pressure for academics to get into print, and the linking of tenure and promotion of academics to publication, has led to a variety of abuses of the system. Singer (1989) cited cases of gross malpractice, for example, where researchers fraudulently claim to have made a new discovery or fabricate research findings. Most ethical violations, however, are less severe but nevertheless significant. As Berardo pointed out:

> Upward mobility (promotion, tenure, recognition, awards, etc.) is facilitated by getting one's name on many publications, and especially if one appears as the single or first name author. Sometimes this leads to having one's name on an article even though the person hasn't written any of it or whose contribution to its composition has been minimal....A related but more insidious pattern is for a the major professor to insist, sometimes subtly and other times bluntly, that graduate students include their names on any publications derived from theses or dissertations completed under their supervision. Such incidents clearly represent violations of the moral and ethical norms which represent the ethos of science. (Berardo, 1989, p. 126)

The issue of intellectual property rights, that is, who owns the ideas, concepts, theories, experimental data, fact and opinions in research articles and reports, has been raised in two contexts. First, electronic publication has been perceived as providing greater possibilities for plagiarism—technically it is relatively simple to cut and paste someone else's text into one's own. The second context involving intellectual property rights of researchers concerns the relationship between government and/or research sponsors (or purchasers), and researchers. A recent concern in the UK has been how to resist pressure on journal editors from government representatives wanting to "pull" papers which are critical of government policy, despite the fact that the papers have satisfactorily scaled all peer review and editorial hurdles. At a time when many academics are exhorted to seek research funding from a range of sources, the UK researcher Nigel Norris (1995, p. 274) draws attention to related problems when government departments sponsor research to support "their strategic objectives and continuing responsibilities." The research community is caught between a rock and a hard place. It needs both to remain "true" to professional standards yet at the same time, avoid being seen as overly critical of sponsors, governments or policies.

One solution to this predicament is not to sign up to such contracts, but there may be good reasons why researchers have little choice; for example, because work will be provided for temporary researchers or the university demands that they gain external funding for research. A strategy evolved to deal with such situations, therefore, has been to develop a code of ethics to be adopted by all partners in a research enterprise which will allow the negotiation of research practice boundaries. Ethical guidelines published by the British Educational Research Association (BERA) which could form the basis of such a code, include the following stipulations regarding academic writing and publication:

- Educational researchers should aim to avoid fabrication, falsification, or misrepresentation of evidence, data, findings, conclusions.
- Educational researchers should aim to report their findings to all relevant stakeholders and so refrain from keeping secret or selectively communicating their findings.
- Educational researchers should communicate their findings and the practical significance of their research in clear, straightforward, and appropriate language to relevant research populations, institutional representatives, and other stakeholders.
- Educational researchers should remain free to interpret and publish their findings without censorship or approval from individuals or organisations, including sponsors, funding agencies, participants, colleagues, supervisors or administrators… (BERA, 1992, 1&2).

Has the academic journal a future?

A key question raised in previous sections of this paper is the extent to which current and future academic cultures and publishing practices might be made more equitable and inclusive. Knowledge of the origins and current state of academic publishing, and debates concerning publishing as a performance indicator and as a site of struggle over power and knowledge as discussed in this paper, suggest that getting a paper published in an academic journal is not nearly as straightforwardly about "good scholarship" as it might at first seem. The impact of technology, literary practices, discourse communities and the power over academic knowledge of like-minded "experts," are all important to our understanding of how academic journals work. The heightened tension in recent years between their utilisation as disseminators of scientific knowledge and as accreditors of scholarship is another factor for consideration.

How can present day academic journals be understood by those aiming to boost their publications count or for beginning researchers or for the wider society which hopes to benefit from its investment in research? Is this the system that we want or need? Does it have to be so unfair? Does electronic publishing offer greater or fewer possibilities for widening academic access and participation to hitherto excluded groups? Some countries, for example Sweden, have not yet succumbed to the academic "publish or perish" ethic so prominent in the US. However, sexism in refereeing practices exposed in a recent study of allocation of research council funding in Sweden (Wennerås & Wold A.,1997) suggests that even in more equity conscious environments, academics, consciously or unconsciously, discriminate in what counts as "excellence" and "scholarship." What are the alternatives to current systems of research evaluation and review?

Briefly there seem to be three main future scenarios:

1. *Stasis*—keeping the system as it is, defending existing cultures of excellence, seeking to impose conventional publishing practices on web-based journals, resisting change;
2. *Deregulation*—reduction of publishing controls, access to technology paramount, a web publishing free-for- all, decline and eventual elimination of the paper journal (while other means are found for evaluating research);
3. *Reform*—comprehensive review of the system, fusing of dual systems of paper and electronic journals, preservation of some form of peer review and quality assurance but re-designed to enhance openness and equity, thinking creatively about how to encourage production, dissemination and exchange of academic knowledge across a variety of communication media, and so on. The Knowledge Exchange Model (KEM) for scholarly publishing proposed by Willinsky (2000) is one step in this direction.

Most academics (apart from Internet specialists and university librarians) seem stuck in the statis scenario, fearing deregulation but unwilling or unable to attempt reform. Reform, nevertheless seems the most promising option, but will need a certain level of conscious attention and commitment for those involved. Editors and referees will need to reflect on the fairness both of their policy regarding acceptance and rejection of papers, and the modes of publication available and appropriate for their present and future readership. University administrators and appointment panels will need to develop more refined and fairer ways of judging research quality, to include, perhaps perusing examples of researchers' work, as in Sweden. Publishers and librarians might work more closely together to see whether a system can be developed which serves both university and market interests. And web-based journal editors will need to develop practices that encourage genuine access and openness rather than merely favouring the privileged academic "nerd" as in the past.

References

Advisory Board for the Research Councils (ABRC) (1990), *Peer Review: a report to the Advisory Board for the Research Councils from the Working Group on peer review*, ABRC

American Educational Research Association (AERA), (undated), List of Minority Scholars to serve as Reviewers, Chairs, Discussants, and on Editorial Boards, unpublished paper

Apple M., (ed.) (1982) *Cultural and Economic Reproduction in Education*, London: Routledge

Association of Research Libraries (1998), *Directory of Electronic Journals, Newsletters and Academic Discussion Lists (7th Edition)*, [on-line] Available: http://www.arl.org:591/index.html

Bakanic V., Simon R. J., and McPhail C., (1987) 'The Manuscript Review and Decision-Making Process,' *American Sociological Review*, 52, 631-642

Bazerman C., (1983), *Reporting the Experiment: changing accounts of scientific doings in the "Philosophical Transactions of the Royal Society 1665- 1800"*, mimeo

Berardo F. M., (1989), Scientific Norms and Research Publication: Issues and Professional Ethics, *Sociological Inquiry*, 59, 3, 119-140

British Educational Research Association (BERA) (1992), *Ethical Guidelines for Educational Research*, Edinburgh, BERA/SCRE

Burbules N. C. & C. Bruce B. C., (1995) "This is not a paper." *Educational Researcher, 24, 8*, 12-18

Clark R. & Ivanic R. (1997) *The Politics of Writing*, London, Routledge

Colman A. M., Garner A. B., and Jolly S., (1992), Research Performance of United Kingdom University Psychology Departments, *Studies in Higher Education*, 17, 1, 97- 103

Eisenstein E. L., (1979), *The Printing Press as an Agent of Change: volumes 1 & 2*, Cambridge, Cambridge University Press

Field J., Lovell T. and Weller P., (1991), Citation Counts and Research Quality in Continuing Education: a cautionary note, *Journal of Further and Higher Education, 15*, 2, 47-53

Foucault, M., (1980) *Power/Knowledge: Selected Interviews and other Writings 1972-1977*, London, Harvester Wheatsheaf.

Furnham A. F., (1990), Quantifying Quality: an arguments in favour of citation counts, *Journal of Further and Higher Education, 14*, 2, 105-110

Garfield E. (1990). How ISI Selects Journals for Coverage: Quantitative and Qualitative Considerations, *Current Comments*, 22, May 28, 5-13

Glass G.V (1999). A new day in how scholars communicate. *Current Issues in Education*, [on-line] 2, 1. Available: http://cie.ed.asu/volumer/number2/index.hmtl

Harnad S. & Hemus M. (1997). All or None: No stable Hybrid or Half-Way Solutions to Launching the Learned Periodical Literature into the PostGutenberg Galaxy, in: (ed.) Butterworth I., *The Impact of Electronic Publishing on the Academic Community*, London, Portland Press, 18-27

Hanish C., Horan J. J., Keen B., & Clark G., (1998) A note on the empirical futility of labor-intensive scoring permutations for assessing scholarly productivity: implications for research, promotion/tenure, and mentoring, *Education Policy Analysis Archives, 6*, 15

Helmreich R. L., Spence J. T., Beane W. E., Lucker G. W. and Matthews K. A. (1980), Making it in academic psychology: demographic and personality correlates of attainment, *Journal of Personality and Social Psychology*, 39, 5, 896-908

Hillage J., Pearson R., Anderson A. & Tamkin P., (1998), *Excellence in Research on Schools*, Research Report RR74, Norwich, HMSO

International Council for Science (ICSU), (1998) *Economics, real costs and benefits of electronic publishing in science – a technical study,* proceedings of ICSU press workshop, Keble College, Oxford, 31 March-2 April

McDermott P., (1994), *Politics and Scholarship: feminist academic journals and the production of knowledge*, Urbana & Chicago, University of Illinois Press

Norris N., (1995), Contracts, control and evaluation, *Journal of Education Policy*, 1, 271-285

Over R., (1982), Research productivity and impact on male and female psychologists, *American Psychologist*, 37, 1, 24-31

Pew Higher Education Roundtable (PHER), (1998), To Publish or Perish, *Policy Perspectives*, 4, 4, 1-12, March

Rambler M., (1999), A New Solution to the Journals Crisis, *The Journal of Electronic Publishing*, 4, 3, 1-6

Singer B. D., (1989), The Criterial Crisis of the Academic World, *Sociological Enquiry*, 39, 2, 127-141

Steig M. F., (1986), *Origins and Development of Scholarly Historical Periodicals*, Alabama, University of Alabama Press

Swales J., (1990) *Genre Analysis: English in Academic and Research Settings*, Cambridge, Cambridge University Press

Townsend B.K., (1993), Feminist Scholarship in Core Higher Education Journals, *The Review of Higher Education*, 17, 1, 21-41

University of Austin Collection and Information Resource Division (1998) Scholarly Communications in Libraries, [on-line]. Available: http://ironwood.lib.utexas.edu.cird/Issues/scholar.html

Van Scoy H. & Oakland T., (1991), Minority Group Literature in Psychology and Education Journals 1952-1973), *Journal of Black Studies*, 22, 2, 301-310

Vrasides C., (2000) Promises of Electronic Forms of Data Representation and Scholarly Publication, *Teachers College Record*, no. 10546

Waaijers L., (1997), Opinion paper: towards a new system of scholarly communication, *Interlending & Document Supply*, 25, 2, 77-8

Ward K. B., Gast J. and Grant L., (1992) Visibility and dissemination of women's and men's sociological scholarship, *Social Problems*, 39, 3, 291-298

Weiner G., (1998), Here a Little, There a Little: equal opportunities policies in higher education in the UK, *Studies in Higher Education*, 23, 3, 321-333

Wennerås C. and Wold A. (1997), Nepotism and sexism in peer review, *Nature*, 387, 341-4

Willinsky J, (2000) Proposing a Knowledge Exchange Model for Scholarly Publishing, *Current Issues in Education*, [on-line] 3, 6. Available: http://cie.ed.asu.edu/volume3/number6/

Young A. E., (1996), The future of surgical journals in the electronic publishing era, *British Journal of Surgery*, 83, 289-290

Zuckermann H., and Merton R. K., (1971), Patterns of Evaluation in Science, *Minerva*, 9, 66-100

About the Author

Gaby Weiner Gaby Weiner is professor of teacher education and research at Umeå University, Sweden. She moved there from a similar post at South Bank University, London in 1998. She has written and edited a number of books and reports on feminism, equity and social justice in education. She has also held a number of positions on various academic journals including journal editor, member of editorial board, editor of special issue, reviewer and of course has also contributed as author.

Correspondence can be sent to:
Gaby Weiner
Teacher Education
Umeå University
S 901 87 Umeå
Sweden
Ttelephone +46 (0) 90 786 7185
fax: +46 (0) 90 786 6671
Email: gaby.weiner@educ.umu.se

http://www.educ.umu.se/~gaby

Copyright 2001 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-

9644). The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Sherman Dorn
University of South Florida

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
hmwkhelp@scott.net

Thomas F. Green
Syracuse University

Alison I. Griffith
York University

Arlen Gullickson
Western Michigan University

Ernest R. House
University of Colorado

Aimee Howley
Ohio University

Craig B. Howley
Appalachia Educational Laboratory

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Benjamin Levin
University of Manitoba

Thomas Mauhs-Pugh
Green Mountain College

Dewayne Matthews
Western Interstate Commission for Higher Education

William McInerney
Purdue University

Mary McKeown-Moak
MGT of America (Austin, TX)

Les McLean
University of Toronto

Susan Bobbitt Nolen
University of Washington

Anne L. Pemberton
apembert@pen.k12.va.us

Hugh G. Petrie
SUNY Buffalo

Richard C. Richardson
New York University

Anthony G. Rud Jr.
Purdue University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Jay D. Scribner
University of Texas at Austin

Michael Scriven
scriven@aol.com

Robert E. Stake
University of Illinois—UC

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)

J. Félix Angulo Rasco (Spain)

Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

other vols. | abstracts | editors | board | submit | comment | subscribe | search

This article has been retrieved 220 times since March 27, 2001

Education Policy Analysis Archives

Volume 9 Number 10 **March 27, 2001** **ISSN 1068-2341**

A peer-reviewed scholarly journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2001, the **EDUCATION POLICY ANALYSIS ARCHIVES.**
Permission is hereby granted to copy any article
if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Japanese EFL Teachers' Perceptions of Communicative, Audiolingual and Yakudoku Activities: The Plan Versus the Reality

Greta Gorsuch
Texas Tech University

Abstract

In recent years, the learning of English as a Foreign Language in Japanese high schools has become the focus of new educational policies applied at the national level. One of these is The Course of Study issue by the Ministry of Education, in which teachers are, for the first time in a long series of curriculum guidelines, adjured to develop students' "positive attitudes towards communicating in English." Another is the JET program, which has put thousands of native English speaking assistant language teachers (ALTs) into Japanese secondary classrooms for the purpose of team teaching with Japanese teachers. Data resulting from a survey project of 876 Japanese high school English teachers was used to provide empirical evidence of teachers' levels of approval of communicative, audiolingual and traditional (*yakudoku*) activities. Teachers were also asked to rate the strengths of a variety of influences on their instruction, including university entrance exams, and pre- and in-service teacher education programs. Teachers' perceptions of both activities and instructional influences were examined in light of teachers' length of career, type of school (private versus public, academic versus vocational), and level of contact with an ALT. The data revealed the complexities of imposing broad, national educational policies on a diverse group of teachers, and in an educational culture which likely precludes teachers' use of communicative activities.

Introduction

In recent years, the teaching of English as a Foreign Language in Japanese secondary schools has become the focus of a variety of new educational policies applied at the national level. In 1989, the Ministry of Education issued a new set of curriculum guidelines and course descriptions for the instruction of English in high schools, called *The Course of Study* (Ministry of Education, Science, and Culture, 1992). For the first time, descriptions for the mainstream, four skills English I and II courses in the new *Course of Study* included the startling injunction that high school teachers were to instill a "positive attitude towards communicating in English" in their students (McConnell, 1995).

Another major change in foreign language education policy in secondary schools applied at the national level was the 1987 advent of the JET program, which brought native English speaking "assistant language teachers" (ALTs) into Japanese junior and senior high school English classes (McConnell, 1995; Wada & Cominos, 1994). The purpose of the JET program was to "provide increased opportunities for interaction in the schools between [ALTs] and Japanese teachers of foreign languages," and by extension, promote the teaching of communicative English (Wada & Cominos, 1994: 1). The JET program is well endowed, with an annual operating budget of US$222,000,000 (McConnell, 1995). The JET program is currently in its twelfth year, and employs 5,361 ALTs from numerous countries ("JET program," 1998). Given the conservative leanings of the Japanese education sector (Lincicome, 1993), these two policies are radical.

However, there are several obvious aspects of the Japanese high school educational culture that work against teachers' acceptance of activities designed to promote students' communicative abilities (McConnell, 1995), implying a mismatch between this politically inspired plan and the realities of Japanese high school EFL education. Further, it is not even clear what Japanese high school English teachers believe about communicative activities. No empirical research on teachers' perceptions based on a generalizable sample has been done, even though *The Course of Study* has been in force in the majority of Japanese high schools since 1992. Observers note that the beliefs of the teachers have not have been taken into account in *The Course of Study* (LoCastro, 1996; Pomatti, 1996; Wada, 1994). There is evidence of this in the JET program as well. According to McConnell (1995), the decision to request ALTs for schools is often made at the prefectural level for political reasons. At the local level then, the day-to-day supervision of ALTs is often left to Japanese teachers of English, who resent the extra workload (Gillis-Furutaka, 1994; McConnell, 1995; Uehara, 1992). The traditional style of reform done by the Ministry of Education is well described by Markee's notion of the center-periphery model of innovation diffusion, in which teachers "merely implement the decisions that are handed down to them" (1997: 63).

This lack of regard for teachers' beliefs about language teaching may be a fatal omission. In contexts in which educational innovations are being implemented, teachers' attitudes take on tremendous importance. Teachers' attitudes and beliefs are the single strongest guiding influence on teachers' instruction (Cuban, 1993; Doyle, 1992; Fang, 1996; Freeman, 1989, 1998; Reynolds & Saunders, 1987; Thompson, 1984).

This article reports Japanese high school English teachers' approval of communicative and non-communicative activities through empirical data resulting from a recent nationwide survey of 876 Japanese EFL high school teachers in nine randomly selected prefectures. The article also describes teachers' perceptions of the circumstances in which they operate, and discusses what effects these circumstances likely have on teachers' approval of communicative activities. This juxtaposition of attitudes and circumstances is suggested by Ajzen (1988), who was concerned about the links between personal attitudes, intentions, circumstances, and personal action; and Markee (1997), who was concerned about the effects of an educational culture on teachers' acceptance of a language education innovation.The presentation and discussion of the data will be used to characterize, from the teachers' point of view, the current state of Japanese EFL education in high schools during a period of time in which sweeping, nationally applied policies have been instituted.

Understanding Teachers' Attitudes: Limitations

Because this study explores teachers' attitudes towards various types of instruction, it is necessary to clarify the relationship between teacher attitudes and actual behavior. For this purpose, Ajzen's model (1988) was adopted. Use of Ajzen's model in EFL/ESL research contexts has been reported in Kennedy and Kennedy (1996). According to Ajzen, an attitude is a person's "evaluative reaction" to some object of interest (1988, p. 23). Ajzen suggested that attitudes then "predispose" the person to creating a cognitive response (a belief) about the object, and a potential to act on the object (an intention). However, positive attitudes towards communicative activities and even positive intentions to do them in the classroom may be influenced by what Ajzen called "subjective norms" and "perceived behavioral control" (p. 133). Ajzen defined "subjective norms" as an influence on intentions arising from a person's "perception of social pressure to perform or not perform the behavior under consideration" (p. 117). Thus, for Japanese high school English teachers, sources of subjective norms would be their students, or colleagues.

Ajzen defined "perceived behavioral control" as "the extent to which people have the required opportunities and resources" to do something (p. 127). Thus, teachers may be hindered in doing communicative activities by "internal" and external" factors of perceived behavioral control (pp. 128-130). Examples for Japanese high school English teachers would be adequate training in communicative methodologies, or textbooks that aided them in creating communicative activities. According to Ajzen's model, then, teachers' attitudes may not be predictive of their behavior. Even though they say they approve of particular types of activities, they may not actually do them in their classrooms. Thus, any data on teachers' attitudes must be interpreted carefully in terms of the realities of teachers' every day work.

The Realities of Japanese High School English Education

There are several aspects of current Japanese high school English education which constitute potential impediments to teachers' acceptance of communicative activities, and thus, the policies of Japanese educational authorities. These are: *yakudoku*, an entrenched traditional method of instruction; high stakes university entrance exams, and inadequate pre- and in-service teacher education programs.

Yakudoku, a traditional method of foreign language instruction, focuses almost exclusively on the translation of English literary texts into Japanese, and direct grammatical instruction in Japanese (Bamford, 1993; Bryant, 1956; Gorsuch, 1998; Henrichsen, 1989; Hino, 1988; Law, 1995). *Yakudoku* has been characterized as an impediment to earlier efforts to change EFL instruction (Henrichsen, 1989, p. 104). In two *yakudoku* classrooms, Gorsuch (1998) observed strongly teacher-centered instruction focused largely on the translation of a difficult English text into Japanese. Both teachers in the study reported that they did not ask the students to produce their own original spoken or written English utterances or sentences, because it would be too "difficult" for students. Clearly, students' abilities to communicate in English could not be developed in such classrooms, in that one of the cornerstones of communicative activities is to create semi-realistic situations in which students can express intended meanings in the second language (Hatch, 1992; Richards & Rodgers, 1986; Terrell, Egasse & Voge, 1982).

There are historical reasons why *yakudoku* remains firmly in place. In postwar Japan during the late 1940s and early 1950s, English language education in secondary schools was marked by a real shortage of English teachers who could speak English and who had sound pedagogical training (Henrichsen, 1989). As a result of post-war teacher education policies designed to quickly increase the number of certified teachers in all fields, large numbers of college graduates who were not proficient in spoken English were made English teachers at secondary schools as a "stop gap measure" (p. 163). Such teachers likely used *yakudoku*, because this is what they knew, and did not have to speak English in order to teach it, a trend which continues today (Kawakami, 1993; Pomatti, 1996; Wakabayashi, 1987).

University entrance exams in Japan are high stakes, and affect the lives of Japanese high school students in many school settings. Many observers have noted strong effects of university entrance exams on classroom instruction in Japan (Eckstein & Noah, 1989; National Institute for Educational Research, 1991; Rohlen, 1983), including English language instruction (Brown & Yamashita, 1995a, 1995b; Gorsuch, 1998; Hildebrandt & Giles, 1983; Kawakami, 1993; Kodaira, 1996; Koike & Tanaka, 1995; Law, 1994, 1995; Miller, 1998; Yukawa, 1994) and on teachers' attitudes towards communicative activities (Gorsuch, 1999a). Reportedly, Japanese high school English teachers feel they are expected to prepare students for university entrance exams by having students translate English passages into Japanese, taking vocabulary quizzes, and focusing their instruction on developing students' linguistic knowledge at the expense of linguistic skills (Law, 1995; Miller, 1998). Many students at academic high schools seem to believe that the purpose of high school English education is university exam preparation (Kodaira, 1996; McConnell, 1995; Pomatti, 1996). Students may influence teachers' instruction through their expectations that teachers are supposed to prepare them for the exams, a phenomenon noted in Japan (Gorsuch, 1999a; Hildebrandt & Giles, 1983), and in other contexts in which high stakes tests are in place (MacDonald & Rogan, 1990; Madaus, 1988; Morris, 1985).

Inadequate pre-service teacher education programs are a third impediment to teachers' acceptance of activities designed to develop students' communicative skills. Current EFL pre-service teacher education programs lack vision and depth of instruction in teaching methodology, and do not provide sufficient teaching practica experiences (Kawakami, 1993; Kizuka, 1997). Many would-be teachers get teaching certificates from universities that do not have an education faculty. Such programs may have little actual interest in teacher preparation (Kizuka, 1997; Kobayashi, 1993). In these programs for EFL teachers at "course approved" universities, would-be teachers need only take a minimum numbers of courses related to English, such as English literature or linguistics. They do not get enough courses which bridge "English language theory and practice" (Kizuka, 1997; National Institute of Educational Research, 1989). The result is a pre-service teacher education system that is inadequate to the task of supporting the development of fundamental changes in instruction implied by policies presented in *The Course of Study* and the presence of ALTs in high schools.

Inadequate in-service teacher education programs are a fourth impediment. On the face of it, it does not seem likely that Japanese in-service programs can produce teachers who have the tools to analyze and change their own teaching, as proposed by Combs (1989), Lortie (1975), and Kanu (1996). Government mandated in-service teacher education in Japan consists of first year induction for new teachers, and very limited in-service courses for experienced teachers. Responsibility for the planning and execution of these programs along Ministry of Education guidelines is left in the hands of prefectural and municipal Boards of Education (Kobayashi, 1993). This has two implications. First, in- service teacher education varies widely in frequency and content from prefecture to prefecture. And second, first year induction and in-service programs are generally provided for public high school teachers, but not for private high school teachers.

"Instructional technique" training for new high school English teachers in Kyoto consists of thirty days of "TEFL training" (Gillis-Furutaka, 1994, p. 34). In Fukui Prefecture, new English teachers at public schools have their teaching observed once by a "High School English Teacher's Consultant," who gives the new teacher "feedback and guidance." In addition, new teachers must undergo a two day seminar in which teachers "learn about game and activity design, motivational strategies, and teaching communicatively" (male Japanese prefectural English faculty in-service program coordinator, personal communication, December 4, 1997).

Public high school English teachers are also required to undergo limited in-service training at later points in their careers. In- service programs can potentially promote the use of communicative activities in Japanese classrooms among senior teachers who may not have had the opportunity to receive training otherwise, and who are "farther away" from their university pre-service training than junior teachers. Indeed, Cohen and Spillane (1992) note that teachers' length of career can influence their attitudes towards instruction. In-service training, if effective, may change senior teachers' attitudes.

Unfortunately, at least one observer, a high school EFL teacher herself, questioned the quality of board of education sponsored in-service education programs, and noted that such programs are offered only for short periods of time (Okada, 1997). Data provided by teaching consultants in Fukui, Nagano, Shizuoka, and Yamaguchi prefectures suggested programs that run from one to three days. The brevity of in-service training for Japanese teachers runs counter to the suggestions of Cohen and Spillane (1992) and MacDonald and Rogan (1990), who stated that effective in-service teacher education should be extended for long periods of time, and conducted while teachers continue their usual teaching schedule.

Finally, due to budget constraints, some prefectures may not offer any specialized EFL in-service teacher education, as in the case of Toyama Prefecture, which discontinued their "English Teacher's Workshop" in 1997 (male Japanese prefectural English faculty in-service program coordinator, personal communication, February 25, 1998). It is apparent that specialized in-service teacher education for EFL teachers is not uniform at the national level. Data from this study may indicate whether teachers' length of career has an effect on their approval of communicative, or other activities, and whether teachers at different stages in their career report that participation in in-service programs influences their instruction.

Diversity in Japanese High School Education

The Japanese high school education system is surprisingly diverse, and *The Course of Study*, a broad national policy, and the JET program, a national level program, are being applied to it. In the research project used to generate the data for this article, teachers at both public and private academic and public vocational and night high schools were surveyed, in order for the data to be generalizable to the population of high school English teachers in Japan. Combined teachers' lists for the nine prefectures revealed that Japanese English teachers at public vocational schools constituted a sizable minority, 783 (12.7%) of all 6,167 teachers in the nine prefectures. Private high school English teachers accounted for 21.8% (1,345)(Gorsuch, 1999a).

From the prefectural teachers' lists, it is apparent these high schools are located in urban areas, and are university-preparation oriented. There is essentially no literature extant focusing on EFL instruction in private academic high schools as specific contexts. There is more literature extant on public vocational and night high schools, although still virtually nothing on EFL programs and teachers specifically. Unfortunately, what there is describes a system of schools which currently have no clear purpose, and where the students have been labeled "low ability." While vocational education at the upper secondary level has been historically intended to fill the labor needs of commerce and industry, vocational and night high schools later became the territory of students who could not successfully compete for admission into colleges or universities (Cantor, 1985; James & Benjamin, 1988). Of direct relevance to high school teachers, Cantor stated "vocational courses find it difficult to recruit good, well qualified teachers" and "both teachers and students suffer from low morale" (p. 71).

James and Benjamin (1988) painted an equally stark picture, suggesting that the Ministry of Education creates guidelines (*The Course of Study*) that keep high school curricula "hard" and fast paced. The guidelines thus act as a screening mechanism to place high school age students in secondary schools appropriate to their academic abilities, as defined by their ability to score well on examinations. The effect of applying a difficult, unitary set of guidelines on a whole population of students with varying abilities in test taking is that high schools in which "low ability" students are concentrated "are given little leeway to address the needs of these students" (39). This may also be true for EFL teachers in vocational high school settings. The data presented in this article may indicate whether such teachers constitute a unique group which responds to the needs of a specific group of students. The data may also indicate whether *The Course of Study* is really applicable to students in vocational and night high schools.

Assistant Language Teachers: The JET Program

The overt purpose of the JET program is to have the assistant language teachers (ALTs) and Japanese teachers of English (JTEs) interact in English and raise JTEs' awareness of English as a communicative medium (Wada & Cominos, 1994b: 1). As such, the JET program offers a powerful potential for instructional change among Japanese teachers of English. Yukawa (1992, 1994) documented changes in the teaching of a male JTE at a high school as a result of team teaching with an ALT. Generally, the JTE stopped using the traditional *yakudoku* translation method and began using communicative methods in class. When the JTE and ALT's teaching relationship ended, however, Yukawa found that the JTE reverted back to teaching in traditional ways. It is possible that the JTE, without the support of the ALT, "disconfirmed" his previous decision to use an educational innovation, in this case, communicative activities (Markee, 1997). Futher research on the persistence of the effects of ALTs on JTEs' instruction seems in order.

It should be noted that team teaching with ALTs is not universally available, or applied. ALTs in the JET program are sent only to schools which formally request them (male Ministry of Education JET functionary, personal communication, September 26, 1997). This means that teachers in some prefectures have more opportunities to teach with ALTs than in others. For example, heavily populated Kanagawa Prefecture has 62 English speaking ALTs in the JET program, while less populous Shizuoka Prefecture has 152 (Ministry of Education, 1997). In addition, schools schedule ALTs for classes in quite different ways, with some schools sending ALTs to a new school every day ("one-shot visits"), to schools that have JTEs and ALTs maintain a regular thrice weekly team teaching schedule in one classroom.

Purpose/Research Questions

The Ministry of Education *Course of Study* has been applied at a national level to Japanese high school EFL teachers at different stages in their careers in very different types of schools, and with variable access to ALTs. It is important to document teachers' responses to the communicative ethos of *The Course of Study* in light of these three variables, and to learn more about their attitudes towards activities associated with other language learning approaches known to be in use in Japan. The research questions are:

> What teaching activities associated with communicative,

> audiolingual, and *yakudoku* approaches to foreign language instruction will Japanese high school English teachers report as being appropriate or not appropriate for English I and II courses? Will teachers' responses differ according to teachers' length of career, type of school, or level of involvement with an ALT?

In addition to documenting teachers' attitudes towards various language learning activities, it is necessary to document teachers' perceived circumstances. Elements of teachers' circumstances would include: teachers' perceptions of the strength of influence of university entrance exams, students' expectations, colleagues' expectations, pre- and in-service teacher education programs, etc. (For a full description of postulated influences in teachers' instruction see Cohen & Spillane, 1992; and Gorsuch, 1999a). In order to compare these data effectively with the results of research question #1, teachers' responses will also be examined in the light of the three variables of teachers' length of career, type of school, and level of involvement with an ALT.

> What influences on instruction will Japanese high school English teachers report as being strong or weak? Will teachers' responses differ according to teachers' length of career, type of school, or level of involvement with an ALT?

Method

Participants

The participants for this study were 876 Japanese high school English teachers at public academic, public vocational, and private academic high schools in nine randomly selected prefectures (Fukui, Kanagawa, Nagano, Saga, Shizuoka, Tokushima, Toyama, Yamagata, and Yamaguchi). Teachers' names were sampled using a systematic random sampling procedure from nine teachers' lists obtained from prefectural boards of education, and from high school teachers in the prefectures. The number of 876 represents a 85% return on the target sample size of 1,035. 340 of the respondents were public academic high school teachers, 277 were public vocational and night high school teachers, and 259 were private academic high school teachers.

Materials

The main data collection instrument providing data for this article was a Japanese-language questionnaire (for the English-language version see the Appendix). The questionnaire had four subsections. Subsection A was designed to capture teachers' attitudes towards classroom activities associated with communicative, audiolingual, and *yakudoku* approaches to foreign language instruction. All three approaches are known to be in current use in Japanese high schools. Teachers were asked to respond to twelve activities in terms of their appropriateness for English I and II courses they were currently teaching by circling a score from 1 ("strongly disagree") to 5 ("strongly agree") under each questionnaire item. To develop the construct validity of the items in this section, eight EFL educator panelists (four of them Japanese, four of them native speakers of English) were asked to categorize a list of 30 activities into the three approaches. Only those items which the panelists were able to unanimously categorize were included in the questionnaire.

Subsection B was designed to establish the grouping variables for the study: teachers' length of career, type of school, and level of involvement with ALTs. Teachers responded to the items by checking one category for each item that fit their situations. For length of career (B1), the three categories were 0-8 years of experience, 9-16 years, and 17+ years. For type of school (B2), the categories were public academic high school, public commercial or industrial high school, public night high school, and private academic high school. Teachers' responses to public commercial, industrial, and night high schools were combined and treated as one category (public vocational high schools). For level of involvement with ALTs (B3), the three categories were teaching English I or II with an ALT at least once a week, less than once a week, and not at all. These grouping variables and their categorical breakdowns were suggested by the literature (Cohen & Spillane, 1992) and a pilot survey conducted by the author (Gorsuch, 1999a).

Subsection C provided the researcher with additional information about the teachers, including their educational experiences. Subsection D was designed to capture teachers' perceptions of the strengths of various influences on their instruction in English I and II classes. On seventeen items, teachers were asked to rate their agreement that a given influence influenced their instruction on a scale from 1 to 5, with 1 indicating "strong disagreement" (a weak influence) and 5 indicating "strong agreement" (a strong influence). The items were inspired by Cohen and Spillane's (1992) notion of "instructional guidance," a model designed to enumerate all possible influences acting on teachers' instruction. The items included in the main questionnaire were items that displayed an adequate degree of construct validity through the earlier pilot survey.

The five page questionnaire was mailed out to teachers in the nine prefectures in three successive waves during spring and summer, 1998, about three weeks apart. Included in each of the first wave of questionnaire envelopes were the questionnaire, a postage paid addressed return envelope, and the gift of a pencil. Teachers were not asked to provide their names when returning the questionnaire. Teachers' responses to items were coded and the data were entered into a MacIntosh PowerBook 5300cs computer on a statistical program, *StatView 4.5* (1995). All analyses were conducted using *StatView 4.5*. Questionnaires with missing data were not included in subsequent analyses.

Analyses

Descriptive statistics for all items in questionnaire subsections A (activities) and D (influences on instruction) ($k = 29$) were calculated including means, standard deviations, skewness coefficients, minimum/maximum scores, and modes. Descriptive statistics for each item split by the three grouping variables (teachers' length of service, type of school, level of involvement with ALTs) were also calculated. Factoral ANOVAs were calculated for each of the 29 comparisons per grouping variable with statistical significance set at $p < .0017$ (.05 divided by 29) to check for significant differences in mean scores on subsection A and D items based on teachers' group memberships. Cronbach's alpha was used to estimate the reliability (internal consistency) of subsection A and D items.

Results

Descriptive statistics for Subsection A are in Table 1. They have been reported from highest mean to lowest.

Table 1
Descriptive Statistics for Activities Items

Item	Approach/Skill	Description	Mean	SD	Skew
A12	Communicative Reading	Students unscramble sentences to make a paragraph.	3.893	.759	-1.15
A11	Communicative Reading	Students match pictures to a story.	3.892	.727	-.97
A5	Audio Lingual Listening/Speaking	Choral repetition of minimal pairs.	3.773	.844	-.81
A3	Communicative Listening/Speaking	Information gap.	3.659	.896	-.59
A6	Audio Lingual Listening/Speaking	Students recite memorized sentence patterns.	3.619	.802	-.56
A8	Audio Lingual Listening/Speaking	Students practice memorized dialogs in pairs.	3.579	.828	-.56
A10	*Yakudoku* Reading	Students unscramble an English sentence suggested by a Japanese translation of the sentence.	3.543	.823	-.83

A1	*Yakudoku* Reading	Students translate English text into Japanese for homework.	3.463	.952	-.59
A9	Communicative Listening/Speaking	Opinion gap.	3.376	.939	-.34
A2	Communicative Writing	Students write predictions of the ending of a picture strip story.	3.372	.900	-.49
A7	Communicative Writing	Students write letters to each other.	3.364	.885	-.37
A4	*Yakudoku* Reading	Studentsrecite their Japanese translations in class.	3.080	1.065	-.30

Teachers gave centered responses on the data. The highest mean score (item A12) was 3.893 and the lowest was 3.080 (item A4). Such centered scores above a "3" indicate a very mild approval of all twelve activities presented to teachers. Teachers in general dwelled in the area between "don't know" (3) and "approve" (4), a conservative and cautious place in which to be. All of the items had a negative skew, which indicated that teachers' responses tended to be bunched up towards the upper end of the distribution created by their scores. This, taken with mode of 4 ("approve") on all items, suggests that as a group, teachers responded in quite similar ways on each item.

Relative approval ratings between items associated with communicative, audiolingual, and *yakudoku* approaches were not entirely clear cut, although teachers were less approving of *yakudoku* activities than expected. However, when items were grouped by level of control of teachers over the language used by students, a more unambiguous pattern emerged. The *yakudoku* items (A1, A4, and A10) aside, teachers approved of controlled activities more than they did activities involving student generation of extemporaneous (non-scripted) language. If items were ranked by mean score from 1 (highest mean score) to 12 (lowest mean score), the six "high teacher/language control" items all rank 6 or above (items A11, A12, A3, A5, A6, and A8), indicating higher approval by teachers. The three "low teacher/language control" items (A2, A7, and A9--all of them communicative items) were ranked at 9, 10, and 11, indicating lower approval by teachers.

Descriptive statistics for Subsection D are in Table 2. These are ranked from highest mean score to lowest.

Table 2
Descriptive Statistics for Influences

Item	Description	Mean	SD	Skew	Min/ Max	Mode
D16	Students' English speaking abilities.	4.318	.652	-1.03	1/5	4
D12	Number of students in class.	4.026	.800	-.86	1/5	4
D2	University entrance exams.	3.905	.987	-.94	1/5	4
D15	Students' expectations.	3.855	.770	-.90	1/5	4
D3	Textbook.	3.701	.839	-.80	1/5	4
D17	Teacher's English speaking ability.	3.620	.846	-.57	1/4	4
D6	Teacher's English learning experiences.	3.558	.986	-.78	1/5	4
D7	Colleagues.	3.094	.925	-.30	1/5	3
D11	Locally written syllabus.	2.986	.907	-.19	1/5	3
D1	Monbusho *Course of Study.*	2.961	.927	-.06	1/5	3
D14	Parents' expectations.	2.634	1.00	.18	1/5	2
D5	In-service teacher education.	2.462	1.19	-.51	0/5	3
D4	Pre-service teacher education.	2.379	.956	.29	1/5	2
D13	Assistant language teacher.	1.879	1.88	.20	0/5	0
D8	Principal.	1.782	.840	1.04	1/5	1
D9	Teacher development courses taken privately.	1.401	1.72	.61	0/5	0
D10	Academic organizations.	.587	1.27	2.60	0/5	0

Teachers' responses were more varied and less centered for subsection D items than on subsection A items. The highest mean score was $M = 4.318$ (students' abilities in English) and lowest was $M = .587$ (membership in an academic organization). For whatever reason, teachers saw no reason to restrict their responses to 3 and 4 on the one to five point Lickert scale as they largely had on subsection A items. Negatively skewed items indicated that teachers' responses tended to be concentrated around the upper end of the distribution created by teachers' scores, while positively skewed items indicated that teachers' responses tended to be concentrated around the lower end of the distribution.

The highest mean score items were D16 ($M = 4.318$, mode = 4) (students' abilities in English) and D12 ($M = 4.026$, mode = 4) (class size). Both indicated strongly that teachers felt these influences in their instruction. Both items represent very "local" influences, which would act directly upon the teachers inside their classrooms. The third, fourth, and fifth highest ranked mean scores belonged to items D2 ($M = 3.905$, mode = 4) (university entrance exams), D15 (*M* = 3.855, mode = 4) (students' expectations), D3 (*M* = 3.701, mode = 4) (textbook), all of which indicated still fairly strong perceptions overall that these influenced teachers' instruction. The sixth and seventh highest mean score items D17 (*M* = 3.620, mode = 4) (teachers' English speaking ability) and D6 (*M* = 3.558, mode = 4) (teachers' experiences learning English as students) indicated moderate agreement that these influence teachers' instruction.

Between the sixth and seventh highest ranked mean score items and the eighth, ninth and tenth highest mean scores is a rather large break of nearly half a point, down to items D7 (*M* = 3.094) (colleagues), D11 (*M* = 2.986) (locally written English I and II syllabuses), and D1 (*M* = 2.961) (Ministry of Education *Course of Study*). These three items were very centered (mode = 3), indicating neither agreement nor disagreement that these influence teachers' instruction.

The eleventh, twelfth, and thirteenth highest mean score items were also in a league of their own, numerically. Items D14 (*M* = 2.634, mode = 2) (expectations of students' parents), D5 (*M* = 2.462, mode = 3) (in-service teacher education), and D4 (*M* = 2.379, mode = 2) (pre-service teaching license program) all represented rather "distant" influences, distant either through time or proximity. Teachers' responses indicated mild disagreement with the notion that these influence instruction.

The lowest four mean score items indicated stronger levels of disagreement that the notions expressed in them influence teachers' instruction. These were D13 (*M* = 1.879, mode = 0) (ALTs), D8 (*M* = 1.782, mode = 1) (the principal), D9 (*M* = 1.401, mode = 0) (teaching courses taken privately), and D10 (*M* = .587, mode = 0) (membership in an academic organization).

On the teacher's length of career grouping variable, six mean scores on Subsection A (activities) and D (influences) items were significantly different by group at $p < .0017$. See Table 3.
br>

Table 3
Significantly Different Mean Scores by Teacher's Length of Career

Item	Item Description	Significantly Different Cells	F-Value
A1	*Yakudoku* reading activity	1 (*M*=3.312) vs. 3 (*M*=3.596)	6.43
A3	Communicative information gap activity	1 (*M*=3.821) vs. 3 (*M*=3.524)	7.90
D6	Influence of English learning experiences on instruction	1 (*M*=3.696) vs. 3 (*M*=3.431)	5.319
D7	Influence of colleagues on	1 (*M*=3.263) vs. 3	5.85

	instruction	(*M*=2.975)	
D9	Influence of privately taken teacher development courses on instruction	1 (*M*=1.058) vs. 2 (*M*=1.619)	7.397
D15	Influence of students' expectations on instruction	1 (*M*=3.962) vs. 3 (*M*=3.751)	5.633

There were some significant differences between teachers on the basis of their length of teaching career. The most senior group of teachers with 17+ years of experience were more likely than the most junior teachers (0-8 years) to approve of a traditional *yakudoku* reading activity (A1). The same senior teachers were less likely to approve of a communicative information gap activity than the most junior teachers (A3). In terms of instructional influences, the junior teachers reported being more strongly influenced by their own language learning experiences, colleagues, and the expectations of students than the senior teachers did (D6, D7, D15). Finally, the middle group of teachers with 9-16 years of experience reported being more strongly influenced by teacher development courses they took privately than the junior group of teachers (D9).

On the type of school grouping variable, eleven mean scores on Subsection A (activities) and D (influences) items were significantly different by group at $p < .0017$. See Table 4.

Table 4
Significantly Different Mean Scores by Type of School

Item	Item Description	Significantly Different Cells	F-Value
A1	*Yakudoku* reading activity	2 (*M*=3.3) vs. 3 (*M*=3.564)	6.216
A3	Communicative information gap activity	1 (*M*=3.762) vs. 3 (*M*=3.471)	8.479
A4	*Yakudoku* reading activity	1 (*M*=3.009) vs. 3 (*M*=3.367) 2 (*M*=2.899) vs. 3 (*M*=3.367)	14.595
D2	Influence of entrance exams on instruction	1 (*M*=4.162) vs. 2 (*M*=3.451) 2 (*M*=3.451) vs. 3 (*M*=4.054)	48.427
D5	Influence of in-service EFL teacher education on instruction	1 (*M*=2.724) vs. 3 (*M*=1.977) 2 (*M*=2.596) vs. 3 (*M*=1.977)	33.711

		(*M*=1.977)	
D7	Influence of colleagues on instruction	1 (*M*=3.209) vs. 3 (*M*=2.965)	5.258
D8	Influence of school principal on instruction	1 (*M*=1.674) vs. 3 (*M*=2.058) 2 (*M*=1.657) vs. 3 (*M*=2.058)	20.631
D11	Influence of locally written syllabus on instruction	1 (*M*=3.079) vs. 2 (*M*=2.827)	6.530
D12	Influence of class size on instruction	2 (*M*=4.123) vs. 3 (*M*=3.869)	7.599
D13	Influence of assistant language teacher on instruction	1 (*M*=2.168) vs. 3 (*M*= 985) 2 (*M*=2.361) vs. 3 (*M*=.985)	47.167
D14	Influence of students' parents' expectations on instruction	1 (*M*=2.656) vs. 2 (*M*=2.397) 2 (*M*=2.397) vs. 3 (*M*=2.857)	14.641

Both public vocational high school English teachers and private academic high school English teachers emerged as singular groups, implying that teachers in these groups have quite different priorities. In terms of influences on instruction, public vocational high school teachers indicated that they were less influenced by university entrance exams than both public academic and private academic high school teachers (D2). Public vocational teachers also reported less influence from their English I and II syllabuses than public academic high school teachers (D11). Finally, public vocational teachers reported being less influenced by students' parents' expectations than private academic high school teachers (D12).

The differences that set private academic high school English teachers apart from teachers in the public sector were more numerous, and point to Japanese private academic high schools as being unique environments. In terms of activities, private academic teachers were more approving of traditional *yakudoku* reading activities than public vocational high school teachers (A1) and public academic and vocational teachers combined (A4). However, private academic high school teachers were less approving of a communicative information gap activity than public academic high school teachers were (A3). Perhaps related to private academic high school teachers' attitudes towards activities is the fact that such teachers reported being less influenced by prefectural in-service teacher education programs than both public academic and vocational high school English teachers (D5). This may imply that such public funded in-service programs are simply not available to private high school teachers. If that is the case, then private high school teachers may have fewer opportunities for professional development, and do not learn about activities such as the communicative information gap activity.

In terms of the influence of human agents on instruction, private academic high school teachers reported being less influenced by their colleagues than public academic high school teachers (D7). However, private academic high school teachers reported being more influenced by their school principals than teachers at either public academic or vocational schools (D8). Finally, private academic high school English teachers reported being much less influenced by ALTs in English I and II courses than either public academic or vocational teachers (D13). This can mean two things: First, private high schools may not have ALTs, and second, private high schools may not use ALTs to team teach in their mainstream English I and II courses and are instead assigned to "oral communication" classes which are less widely offered (the latter has been strongly suggested in Gorsuch, 1999a).

On the level of involvement with an ALT grouping variable, only two mean scores on Subsection A (activities) and D (influences) items were significantly different by group at $p < .0017$. See Table 5.

Table 5
Significantly Different Mean Values by Level of Involvement with an ALT

Item	Item Description	Significantly Different Cells	F-Value
A3	Communicative information gap activity	1 (M=3.876) vs. 3 (M=3.518) 2 (M=3.879) vs. 3 (M=3.518)	17.440
D13	Influence of assistant language teacher on instruction	1 (M=3.601) vs. 3 (M=.856) 2 (M=3.327) vs. 3 (M=.856)	380.547

Teachers teaching with ALTs more than once a week, or less than once a week approved of the communicative information gap activity more than teachers with no ALT contact (A3). And, not surprisingly, teachers teaching with ALTs more, or less, than once a week reported being much more influenced by ALTs than teachers with no ALT contact (D13).

Cronbach's alpha internal consistency coefficient for subsection A and D items was only .6878, which was only moderate. Subsection A and D items purportedly measure several different constructs, which will depress internal consistency estimates. In addition, teachers' responses to subsection A items (activities) were very centered (values all around "3"). Such homogeneous values will probably depress internal consistency estimates. In addition to the constructs the researcher intended to measure, there was some measurement error, as indicated by the moderate reliability coefficient.

Discussion

What activities do teachers approve of? The results indicated that teachers have generally positive attitudes towards communicative language teaching (CLT) activities. However, teachers seemed to prefer the more highly controlled, passive skill, CLT activities over CLT activities that called for students to engage in extemporaneous (non-scripted, non-memorized) speech and writing. Teachers' greater preferences for controlled CLT activities were matched by strong preferences for the audiolingual activities, which involved the students' use of memorized speech in pattern practice drills or dialogs. Thus, the teachers seemed to indicate that CLT activities were alright, as long as the teachers could control students' language while using them. The teachers seemed to be responding in a cautious, although positive, way towards communicative activities.

Gorsuch (1998) described the two high school English teachers she observed as being overwhelmingly concerned with student accuracy. There may be perfectly justifiable reasons for teachers' desire for control. Japanese classes typically have at least 40 students in them (Gorsuch, 1998; Kawakami, 1993). With such a large class, it would be easy to "lose control" of students during a communicative speaking activity. In addition, teachers night feel hard pressed to effectively monitor 20 or more pairs of talking students. Yet *The Course of Study* specifically mentions helping students develop a positive attitude towards communication. If students are to do so, they have to be allowed and encouraged to communicate in class. The reasoning behind this is, how can students develop a positive attitude towards communication if they do not actually experience communication? In the end, teachers may have to learn to give up a measure of control over students' use of English, and demand smaller classes.

The communicative information gap activity A3 seemed to be a kind of litmus test for approval or non-approval of CLT activities based on group membership. Teachers who approved of A3 more highly were younger teachers, teachers at public academic high schools, and teachers who had at least some contact with ALTs. Teachers who did not approve of A3 as much were older teachers, teachers at private academic high schools, and teachers with no contact with ALTs. Concerning teachers' length of career, more senior teachers may not approve of A3 because they have been out of pre-service teacher education programs longer than junior teachers. This, coupled with what seems to be a real lack of in- service teacher education programs, and a lack of interest on the part of teachers in taking professional development courses privately or belonging to academic organizations (Table 2) may imply that senior teachers have not had sufficient training to feel comfortable trying out an activity like A3 for themselves.

Most interesting, though, was the greater approval of A3 by teachers teaching at least once a week or less than once a week with

ALTs than teachers not teaching with an ALT at all. Perhaps teachers who have regular contact with ALTs find it easier to model CLT pair work activities for students. It could also be that when an ALT is in the classroom, students expect to do something different than highly controlled language practice. There may also be a link with teachers' self-perception of English speaking skill--in a separate analysis of teachers' self ratings of English speaking skill, it was found that teachers teaching with ALTs at least once a week rated their English speaking skills significantly higher than teachers who had less or no contact with ALTs (Gorsuch, 1999a). Whether a causal factor or not, presence of an ALT is linked with greater approval of A3 and higher self reports of teacher English speaking ability.

There was one difference on teachers' approval of *yakudoku* item A4 due to group membership. Teachers at public academic and vocational high schools were less likely to approve of having students recite their Japanese translations in class than private academic high school teachers. One possible reason is that private academic high school teachers seem to be largely excluded from in-service teacher education offered by prefectural or municipal boards of education, where they may receive training in other methodologies.

Teachers' responses to all of the activity items in the questionnaire were centered around "3" (Table 1). When "significant" differences in level of approval or disapproval are discussed above, such differences were very subtle, sometimes representing half a point or less of difference on a five point scale. This was a disappointing result, yet not altogether unexpected, given the general conservatism of educators in Japan. *The Course of Study* is asking teachers to do something quite new--develop students' communicative abilities--and teachers are responding cautiously, and obviously only within the bounds of their understanding of what both spoken and written communicative activities entail.

What influences teachers? Teachers responded to items in subsection D in non-centered fashion. Perhaps they felt less cautious and constrained when asked to respond to "safer," less ideologically laden, items. Unfortunately, teachers' responses indicated that there were powerful impediments working against their acceptance of CLT activities, such as the strong influences of university entrance exams and students' expectations, and the surprisingly weak influences of pre- and in-service teacher education programs, and privately undertaken courses.

With the exception of the entrance exams item (D2), teachers generally agreed that students' English abilities (D16), class size (D12), students' expectations (D15), the textbook (D3), teachers' English speaking abilities (D17), and teachers' English learning experiences (D6) exerted powerful influence on their instruction. Some of these may prevent teachers from teaching communicatively. It is not surprising that teachers consider their students' abilities to be a crucial factor in planning instruction. No teacher wants to go into a classroom with a lesson plan that is too easy or too difficult for the students. Activities of the first type will bore them, and the second type will stymie and then bore them. Either case implies teachers' loss of control over the class, something Japanese teachers have indicated through their activity preferences as undesirable to them. Unfortunately, Japanese teachers seem to consider communicative activities to be "difficult," even for students in top ranked high schools (Gorsuch, 1998). If other teachers with less able students share this perception, then teachers will likely not use communicative activities, regardless of their cautious approval suggested in this study.

As noted above, class sizes are large (40+). Teachers are likely concerned whether they will be able to control such a large group of students. This perception, coupled the high influence rating teachers gave to the student expectations item (Table 2), gives the feeling that teachers may be very sensitive to losing control of the students by going against students' expectations. Recall the observations of scholars cited earlier that the majority of students expected their English class work to prepare them for entrance exams. In such a climate, teachers are unlikely to feel they can comfortably use communicative activities in class.

In terms of teachers' ratings of the influence of textbooks, current Ministry of Education approved English I and II textbooks largely focus on developing students' intensive reading skills for entrance exam preparation, and do not provide aid to teachers in developing communicative activities (Gorsuch, 1999b). This does not bode well for communicative activities, in that appropriate textbooks are necessary to successful implementation of educational innovations (MacDonald & Rogan, 1990).

There really is no escape from the influence of university entrance exams, apparently. Not only did teachers give exams a high rating, exams make their influence known through students' expectations, and through textbooks. There was one difference on the grouping variable B2 (type of school) on the university entrance exam items, however. Public vocational high school teachers were less likely to report that university entrance exams influenced their instruction than teachers at public and private academic high schools. Vocational public high schools may be the perfect venue in which to introduce programs with genuinely communicative aims. Because teachers (and, possibly the students) in these schools feel less influenced by the need to prepare their students for university entrance exams, teachers could, with concerted help, develop English courses making use of suitable communicative activities. If well designed, such activities can be motivating to students who traditionally have little desire to learn English, especially in the traditional exam preparation oriented way (*yakudoku*). Rather than being seen as the sad realm of students who cannot compete academically in the prevailing educational culture, the public vocational high school sector could be an important venue for meaningful instructional change that can later be adapted to the public and private academic high schools. This view of public vocational high schools is in accord with recent efforts to revitalize vocational high school education in Japan ("Vocational school curriculum urged to include scuba diving," 1998).

Teachers reported "colleagues," "locally written syllabuses," and *The Course of Study* as having a neutral influence on their instruction (Table 3). However, the youngest group of teachers (0-8 years of experience) reported colleagues as being more influential than middle (9-16 years) and senior teachers (17 years years) did. Given junior teachers' newness to teaching in specific contexts, it is not surprising that they need the help of more senior teachers to show them the ropes. Whether this help centers on actual teaching in English classrooms is not known.

Providing yet another argument for the adoption of alternative language programs in public vocational high school, teachers at those schools reported that locally written syllabuses influenced their instruction less than teachers in academic high school contexts did. With students who cannot compete to enter universities, vocational schools are left behind in terms of their locally written syllabuses, which are local tokens of *The Course of Study*. A syllabus may be written, but teachers will not, or cannot follow them, perhaps due to students' low academic interests and abilities.

One of the most distressing findings of this study was the low influence status accorded by all teachers to pre-service, in-service, and privately undertaken teacher education courses (Table 2). Either in-service or private courses are not available to teachers, or teachers do not avail themselves of them. Pre-service courses may simply not be attuned to current and future teachers' needs. These circumstances are a negative indictment of foreign language education in Japan. Without adequate pre-service and continuing teacher education, teachers cannot learn about the theoretical bases of different language learning approaches, nor get guided experiences in using them. In this non-teacher-development climate, it is difficult to see how teachers can realistically try communicative activities. However, there was a ray of hope in that teachers with 9-16 years of teaching experience were more likely to report that privately undertaken teacher education courses were influential than the youngest teachers (Table 3). It may be that these middle-aged teachers represent a group of potential users of communicative activities in that they may have confidence in their teaching seasoned by experience, yet feel they want further knowledge and variation in their working lives. The Ministry of Education and local boards of education may wish to develop more intensive and flexible in-service programs aimed specifically at this group of teachers.

Conclusion

It is clear that there is no one solution to enhancing teachers' approval of the communicative activities called for by *The Course of Study* and the continued presence of ALTs in Japanese high school EFL classrooms. This article has given empirical evidence suggesting that teachers mildly approve of communicative activities, yet the data also suggested there are potent impediments working against teachers actually using such activities in their classrooms. This article has also shown how these impediments may work on teachers, from the teachers' point of view.

It is a time of extraordinary change in Japanese high school EFL education. This article has provided an empirical snapshot of the perceptions of Japanese EFL high school teachers, and how these policy changes may potentially affect them. Needless to say, to track future change, further study aimed at gathering empirical data is needed from a variety of points of view. The author hopes that the Ministry of Education, and particularly high school teachers themselves, will undertake such research and take the results into account when planning future curriculum revisions, teacher education programs, or research projects.

References

Ajzen, I. 1988: *Attitudes, personality, and behavior*. Milton Keynes: Open University Press.

Bamford, J. 1993: Beyond grammar translation. In Wadden, P., editor, *A Handbook for Teaching English at Japanese Colleges and Universities*, Oxford: Oxford University Press, 63-71.

Brown, J.D. & Yamashita, S. 1995a: English language entrance exams at Japanese universities: What do we know about them? *JALT Journal*, 17: 7-30.

Brown, J.D. & Yamashita, S. 1995b: English language entrance examinations at Japanese universities: 1993 and 1994. In J.D. Brown and S. Yamashita, editors, *JALT applied materials: Language testing in Japan*, Tokyo: Japan Association for Language Teaching, 86-100.

Bryant II, W. C. 1956: English language teaching in Japanese schools. *The Modern Language Journal*, 71: 21-48.

Cantor, L. 1985: Vocational education and training: The Japanese approach. *Comparative Education*, 21: 67-76.

Cohen, D.K. & Spillane, J.P. 1992: Policy and practice: The relations between governance and instruction. *Review of Research in Education*, 18: 3-49.

Combs, A.W. 1989: New assumptions for teacher education. *Foreign Language Annals*, 22: 129-134.

Cuban, L. 1993: *How teachers taught: Constancy and change in American classrooms 1880-1990*. New York: Teachers College Press.

Doyle, W. 1992: Curriculum and pedagogy. In Jackson, P.W., editor, *Handbook of research on curriculum* New York: MacMillan Publishing Company, 486-516.

Eckstein, M. & Noah, H. 1989: Forms and functions of secondary school leaving examinations. *Comparative Education Review*, 33: 295-316.

Fang, Z. 1996: A review of research on teacher beliefs and practices. *Educational Research*, 38: 47-65.

Freeman, D. 1989: Teacher training, development and decision making: A model of teaching and related strategies for language teacher education. *TESOL Quarterly*, 23: 27- 45.

Freeman, D. 1998: *Doing teacher research: From inquiry to understanding*. Pacific
Grove: Heinle and Heinle.

Gillis-Furutaka, A. 1994: Pedagogical preparation for JET Programme teachers. In M. Wada and T. Cominos, editors, *Studies in Team Teaching*, Tokyo: Kenkyusha, 29- 41.

Gorsuch, G. 1998: *Yakudoku* EFL instruction in two Japanese high school classrooms: An exploratory study. *JALT Journal*, 20: 6-32.

Gorsuch, G. 1999a: Exploring the relationship between educational policy and instruction in Japanese high school EFL classrooms. Unpublished doctoral dissertation, Graduate School of Education, Temple University, Tokyo.

Gorsuch, G. 1999b: Monbusho approved textbooks in Japanese high school EFL classes: An aid or a hindrance to educational policy innovations? *The Language Teacher 23*(10,
5,7,8-9, 11-15.

Hatch, E. 1992: *Discourse and language education*. Cambridge: Cambridge
University Press.

Henrichsen, L.E. 1989: *Diffusion of innovations in English language teaching: The ELEC effort in Japan, 1956-1968.* NY: Greenwood Press.

Hildebrant, N. & Giles, H. 1983: The Japanese as subordinate language group: Ethnolinguistic identity theory in a foreign language context. *Anthropological Linguistics*, 24: 436-466.

Hino, N. 1988: Yakudoku: Japan's dominant tradition in foreign language learning. *JALT Journal*, 10: 45-55.

James, E. & Benjamin, G. 1988: *Public policy and private education in Japan*. New York: St. Martin's Press.

JET program kicks off 12th year 1998, July 23: *The Daily Yomiuri*: 1.

Kanu, Y. 1996: Educating teachers for the improvement of the quality of basic education in developing countries. *International Journal of Educational Development,* 16: 173-184.

Kawakami, H. 1993: Factors influencing English education in Japanese high schools: A survey of teacher perceptions. Unpublished master's thesis, Brigham Young University.

Kennedy, C. & Kennedy, J. 1996: Teacher attitudes and change implementation. *System,* 24: 351-360.

Kizuka, M. 1997, October: *A consideration of the problems in secondary school pre-service teacher education in Japan*. Paper presented at the meeting of the Japan Association for Language Teaching, Hamamatsu, Japan.

Kobayashi, T. 1993: Japan's teacher education in comparative perspectives. *Peabody Journal of Education,* 68: 4-14.

Kodaira, F. 1996: Why not the "Japanese way"?: The traditional approach of learning English as a foreign language at secondary schools in Japan. Unpublished M.A. thesis, Tokyo, Columbia Teacher's College.

Koike, I. & Tanaka, H. 1995: English in foreign language education policy in Japan: Toward the twenty-first century. *World Englishes,* 14:13-25.

Law, G. 1994: College entrance exams and team teaching in high school English classrooms. In M. Wada and T. Comonos, editors, *Studies in team teaching*, Tokyo: Kenkyusha, 90-102.

Law, G. 1995: Ideologies of English language education in Japan. *JALT Journal,* 17: 213-224.

Lincicome, M. 1993: Nationalism, internationalization, and the dilemma of educational reform in Japan. *Comparative Education Review,* 37: 123-151.

LoCastro, V. 1996: English language education in Japan. In Coleman, H., editor, *Society and the language classroom*, Cambridge: Cambridge University Press, 40-58.

Lortie, D.C. 1975: *Schoolteacher: A sociological study.* Chicago: University of Chicago Press.

MacDonald, M.A. & Rogan, J.M. 1990: Innovation in South African science education (part 2): Factors influencing the introduction of instructional change. *Science Education,* 74: 119-132.

Madaus, G. 1988: The influence of testing on the curriculum. In Singerman, A.J., editor, *Critical issues in the curriculum*, Chicago: University of Chicago Press, 83- 121.

Markee, N. 1997: *Managing curricular innovation.* Cambridge: Cambridge University Press.

McConnell, D.L. 1995: Japan JETs international: Implementing innovations in educational policy. In Montergomery, J.D. and Rondinelli, D.A., editors, *Great policies: Strategic innovations in Asia and the Pacific Basin*, Westport, CT:
Praeger, 75-97.

Miller, T. 1998, June: *Comparing the "fit" between communication-oriented teaching and the ELT landscape in Japan.* Paper presented at the Temple University Japan colloquium on Applied Linguistics, Tokyo, Japan.

Ministry of Education, Science, and Culture 1992: *The course of study for senior high school: Foreign languages (English).* Tokyo: Author.

Ministry of Education, Science, and Culture 1997: *ALTs in the prefectures.* Tokyo: Author.

Morris, P. 1985: Teachers' perceptions of the barriers to the implementation of a pedagogic innovation: A South East Asian case study. *International Review of Education,* 31: 3-17.

National Institute for Educational Research 1989: *Teacher training in Japan* (ERIC Document Reproduction Service No. ED313 360).

National Institute for Educational Research 1991: *Toward formulating goals, aims, and objectives of secondary education for the 21st century.* Tokyo: Author.

Okada, J. 1997, October: *Teacher education programs: Problems and solutions.* Paper presented at the meeting of the Japan Association for Language Teaching, Hamamatsu, Japan.

Pomatti, D. 1996: English language education in the Japanese public schools: Obstacles to a communicative approach and Ministry of Education policies. Unpublished manuscript.

Reynolds, J. & Saunders, M. 1987: Teacher responses to curriculum policy: Beyond the "delivery" metaphor. In Calderheard, J., editor, *Exploring teachers' thinking*, London: Cassell Educational Limited, 195-214.

Richards, J.C. & Rodgers, T.S. 1986: *Approaches and methods in language teaching*. Cambridge: Cambridge University Press.

Rohlen, T. 1983: *Japan's high schools*. Berkeley, CA: University of California Press.

Statview 4.5 (Computer software). 1995: Berkeley, CA: Abacus Concepts.

Terrell, T.D., Egasse, J., & Voge, W. 1982: Techniques for a more natural approach to second language acquisition and learning. In Blair, R.W., editor, *Innovative approaches to language teaching*, Rowley, MA: Newbury House Publishers, Inc., 174-175.

Thompson, A.G. 1984: The relationship of teachers' conceptions of mathematics and mathematics teaching to instructional practice. *Education Studies in Mathematics*, 15: 105-127.

Uehara, S. 1992: AET seido no naka de kuno suru eigo kyoshitachi [Teachers who are annoyed with the JET program]. *The New English Classroom,* 8: 8-10.

Vocational school curriculum urged to include scuba diving. 1998, July 24: *The Daily Yomiuri*, 2.

Wada, M. 1994: Team teaching and the revised course of study. In Wada, M. and Cominos, T., editors, *Studies in team teaching*, Tokyo: Kenkyusha, 7-16.

Wada, M. & Cominos, T. (Eds.). 1994a: *Studies in team teaching.* Tokyo: Kenkyusha.

Wada, M. & Cominos, T. 1994b: Language policy and the JET program. In Wada, M. and Cominos, T., editors, *Studies in team teaching*. Tokyo: Kenkyusha, 1-6.

Wakabayashi, S. 1987: Nihon no Eigo kyoiku no ayumi. *Shin eigo kyoiku*. Tokyo: Sanyusha, 145-146.

Yukawa, E. 1992: Team teaching and changes in teaching routines. *The Language Teacher,* 18: 9, 11, 13.

Yukawa, E. 1994: Team teaching and changes in teaching routines in a

Japanese high school reading classroom. In Wada, M. and Cominos, T., editors, *Studies in team teaching*, Tokyo: Kenkyusha, 42-60.

About the Author

Greta Gorsuch, Ed.D., taught EFL in Japan for fifteen years. Former editor of the The Language Teacher (The Japan Association for Language Teaching) and co-author of the Impact series (Lingual House Publishers), she is now Assistant Professor of Applied Linguistics at Texas Tech University in Lubbock, Texas. Her research interests include: Rasch analysis, second language testing, teacher learning, performance assessment. Contact information: e-mail: greta.gorsuch@ttu.edu, address: Classical and Modern Languages and Literatures, Box 42071, Texas Tech University, Lubbock, TX 79409-2071, U.S.A.

Appendix

Questionnaire (English Version)

This questionnaire is designed for teachers who are currently teaching English I and/or English II. If you are not teaching these courses this year, please give this questionnaire to a colleague who is teaching English I and/or English II this year. Thank you!

Please read the activity descriptions below and write a circle or check in the blank that best describes your level of agreement. Please consider each activity carefully, and let your response reflect your true impression about the appropriateness of the activities for your current English I or II classes. If you choose "5" for example, this means you would be strongly willing to use the activity in your class. If you choose "1", this means, you would not be at all willing to use the activity. Please choose only one response.

Items are rated on a 5-point scale from Strongly Agree to Strongly Disagree with "Don't Know" as the middle option.

A-1. The teacher asks students to translate English phrases or sentences into Japanese as preparation for class. I think the above is an appropriate activity for my English I or English II classes: SA A DK D SD

A-2. The teacher has students look at a page that has a "picture strip story." Students can uncover only one picture at a time. Before uncovering the next picture, the students predict, writing the prediction in English, what will happen in the next picture. Students can then look at the next picture to confirm or disconfirm their predictions. I think the above is an appropriate activity for my English I or English II classes:

A-3. The teacher has the students work face to face in pairs. One student sees a page that has some missing information. The other student sees a different page that has that information. The first student must ask questions in English to the other student to find the missing information. I think the above is an appropriate activity for my English I or English II classes:

A-4. The teacher asks students to translate English phrases or sentences into Japanese in preparation for class. Then in class, the teacher calls on individual students to read their Japanese translation of an English phrase or sentence, and the teacher corrects it if necessary and gives the whole class the correct translation with an explanation. I think the above is an appropriate activity for my English I or English II classes:

A-5. The teacher has students chorally repeat word pairs such as sheep/ship and leave/live. I think the above is an appropriate activity for my English I or English II classes:

A-6. The teacher has students memorize and practice a short English sentence pattern. The teacher then gives the students a one word English cue and has the students chorally say the sentence pattern using the new word. I think the above is an appropriate activity for my English I or English II classes:

A-7. The teacher pairs off students. Then the teacher asks the students to write a letter in English to their partner. I think the above is an appropriate activity for my English I or English II classes:

A-8. The teacher has students memorize an English dialog and then has the students practice the dialog together with a partner. I think the above is an appropriate activity for my English I or English II classes:

A-9. The teacher has pairs or small groups of students ask each other and then answer questions in English about their opinions. I think the above is an appropriate activity for my English I or English II classes:

A-10. Students read a sentence in Japanese, and then see an equivalent English sentence below where the words been scrambled up. The students must then rewrite the English sentence in the correct order suggested by the Japanese sentence. I think the above is an appropriate activity for my English I or English II classes:

A-11. On one page students see a picture. Underneath the picture are several short English stories. Students have to choose which story they think best matches the picture. I think the above is an appropriate activity for my English I or English II classes:

A-12. On a page, students see an English paragraph in which the sentences have been scrambled. The teacher then asks the students to put the sentences into order so the paragraph makes sense. I think the above is an appropriate activity for my English I or English II classes:

A-13. What activity do you feel is most effective for your students in your English I or II class? Please write a brief description here: (Optional)

Please answer the following questions by writing a check next to the most correct answer. Choose only one response.

B-1. How many years have you been teaching in high school?_____ 0-8 years_____ 9-16 years_____ 17+ years

B-2. What kind of high school are you currently teaching in?_____ public academic high school_____ public commercial or industrial high school_____ public night high school_____ private academic school

B-3. Are you currently teaching English I or English II with an ALT (Assistant Language Teacher)?_____ Yes, at least once a week._____ Yes, but less than once a week._____ No, I do not teach English I or English II with an ALT

Please read the sentences below and write a check in the blank that best describes your level of agreement. Choose only once response.

C-1. My English speaking ability is good enough for me to use in class.

C-2. As a student I studied English primarily through translating English stories, essays, or literary works into Japanese.

C-3. I think the pace we have to teach English at my high school is:much too fast____ fast____ about right____ slow____ much too slow____

C-4. The average size of my English I or English II classes is:over 50____ 40-49____ 30-39____ 20-29____ below 19____

Please read the sentences below concerning your current instruction in English I and II classes and write a check in the blank that best describes your level of agreement. Choose only one response.

D-1. The Monbusho guidelines for English I and English II influences my classroom practice.

D-2. College and university entrance exams influence my classroom practice.

D-3. The textbook my students are using influences my classroom practice.

D-4. The teaching license program I completed at university influences my current classroom practice.

D-5. In-service teacher education specifically designed for English teaching offered by my prefectural or municipal board of education influences my classroom practice.
In-service teacher education for English teaching is not available from the Board of Education for me.

D-6. The way I learned English as a student influences my current classroom practice.

D-7. My English teaching colleagues influence my classroom practice.

D -8. The principal at my school influences my classroom practice.

D-9. Teaching courses I have taken privately influence my current classroom practice.
_____ I have not taken teaching courses privately.

D-10. My membership in a private academic organization influences my classroom
_____ I am not a member of an academic organization.

D-11. The English I and English II syllabus used at my school influences my classroom practice.

D-12. The number of students in my English I or II classes influences my classroom practice. (i.e., Would you teach differently if your classes had many students or few students?)

D-13. The ALT I teach English I or II with influences my classroom practice.
_____ I do not currently teach English I or English II with an ALT.

D-14. The expectations of my students' parents influences my classroom practice.

D-15. My students' expectations about how to study English influences my classroom practice.

D-16. My students' abilities in English influences my classroom practice.

D-17. My level of English speaking ability influences my classroom practice.

D-18. What is one influence not listed above that you feel strongly influences your instruction of English I or English II? (Optional)

Copyright 2001 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin
John Covaleskie
Northern Michigan University
Sherman Dorn
University of South Florida
Richard Garlikov
hmwkhelp@scott.net
Alison I. Griffith
York University
Ernest R. House
University of Colorado
Craig B. Howley
Appalachia Educational Laboratory
Daniel Kallós
Umeå University
Thomas Mauhs-Pugh
Green Mountain College
William McInerney
Purdue University
Les McLean
University of Toronto
Anne L. Pemberton
apembert@pen.k12.va.us
Richard C. Richardson
New York University
Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Greg Camilli
Rutgers University
Alan Davis
University of Colorado, Denver
Mark E. Fetler
California Commission on Teacher Credentialing
Thomas F. Green
Syracuse University
Arlen Gullickson
Western Michigan University
Aimee Howley
Ohio University
William Hunter
University of Calgary
Benjamin Levin
University of Manitoba
Dewayne Matthews
Western Interstate Commission for Higher Education
Mary McKeown-Moak
MGT of America (Austin, TX)
Susan Bobbitt Nolen
University of Washington
Hugh G. Petrie
SUNY Buffalo
Anthony G. Rud Jr.
Purdue University
Jay D. Scribner
University of Texas at Austin

for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert E. Stake
University of Illinois—UC

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

other vols. | abstracts | editors | board | submit | comment | subscribe | search

This article has been retrieved 404 times since April 2, 2001

other vols. | abstracts | editors | board | submit | comment | subscribe | search

Education Policy Analysis Archives

Volume 9 Number 11 **April 2, 2001** **ISSN 1068-2341**

A peer-reviewed scholarly journal
Editor: Gene V Glass, College of Education
Arizona State University

Copyright 2001, the **EDUCATION POLICY ANALYSIS ARCHIVES**.
Permission is hereby granted to copy any article
if **EPAA** is credited and copies are not sold.

Articles appearing in **EPAA** are abstracted in the *Current Index to Journals in Education* by the ERIC Clearinghouse on Assessment and Evaluation and are permanently archived in *Resources in Education*.

Constructing Outcomes in Teacher Education: Policy, Practice and Pitfalls

Marilyn Cochran-Smith
Boston College

Abstract

As we enter the twenty-first century, the outcomes, consequences, and results of teacher education have become critical topics in nearly all of the state and national policy debates about teacher preparation and licensure as well as in the development of many of the privately and publicly funded research agendas related to teacher and student learning. In this article, I argue that teacher education reform over the last fifty years has been driven by a series of questions about policy and practice. The question that is currently driving reform and policy in teacher education is what I refer to as "the outcomes question." This question asks how we should conceptualize and define the outcomes of teacher education for teacher learning, professional practice, and student learning, as well as how, by whom, and for what purposes these outcomes should be documented, demonstrated, and/or measured. In this article, I suggest that the outcomes question in teacher education is being conceptualized and constructed in quite different ways depending on the policy, research, and practice contexts in which the question is posed as well as on the political and professional motives of the posers. The article begins with an overview of the policy context, including those reforms and initiatives that have most influenced how outcomes are currently being constructed, debated, and enacted in teacher education. Then I identify and analyze three major "takes" on the outcomes question in teacher education—outcomes as the long-term or general impacts of teacher education, outcomes as teacher candidates' scores on high stakes teacher tests, and outcomes as the professional performances of teacher candidates, particularly their demonstrated ability to influence student learning. For each of these approaches to outcomes, I examine underlying assumptions about teaching and schooling, the evidence and criteria used for evaluation, units of analysis, and consequences for the profession. I point out that how we construct outcomes in teacher education (including how we make the case that some outcomes matter more than others) legitimizes but also undermines particular points of view about the purposes of schooling, the nature of teaching and learning, and the role of teacher education in educational reform. In the second half of the article, I offer critique across the three constructions of outcomes, exploring the possibilities as well as the pitfalls involved in the outcomes debate. In this section, I focus on the tensions between professional consensus and critique, problems with the inputs-outputs metaphor, the need to get social justice onto the outcomes agenda, problems with the characterization of teachers as either saviors or culprits, and the connection of outcomes to educational reform strategies that are either democratic or market-driven.

In public opinion polls of what concerns Americans most, education has ranked higher than the economy, the environment, and even crime (Mosle, 1996). Since 1996, the *New York Times* alone has printed 1,220 articles about teacher quality and 920 articles about teacher testing. And, as the following excerpt from the first Bush-Gore presidential debate indicates, the quality of public schools and of the nation's teaching force has now reached center stage in national politics (not to mention its continued central role in state and local politics):

> Mr. Lehrer (Debate Moderator): All right. So, having heard the two of you, voters have just heard the two of you, what's the difference? What's the choice between the two of you on education?
>
> Mr. Bush: Well the first—first is, the difference is, there is no new accountability measures in Vice President Gore's plan. He says he's for voluntary testing. You can 't have voluntary testing. You must have mandatory testing. You must say that if you receive money, you must show us whether or not children are learning to read and write and add and subtract. That's the difference. You may claim you've got mandatory testing, but you don't. Mr. Vice President. And that is a huge difference. Testing is the cornerstone of reform…
>
> Mr. Gore: Well first of all, I do have mandatory testing. I think the governor may not have heard what I said clearly. The voluntary national test is in addition to the mandatory testing that we will require of states—all schools, all school districts, students themselves and required teacher testing, which goes a step farther than Governor Bush has been willing to go (*New York Times* Archives, 2000).

These comments from then presidential candidates George Bush and Al Gore reflect the current national attention to teacher quality and its frequent identical twin, teacher testing. In the media, in public policy debates, and within the profession of teaching and teacher education itself, there is unprecedented emphasis on accountability, results, and outcomes, or at a fundamental level, what connection the public has a right to expect among teaching, schooling, and student learning.

In this article, I consider these issues by focusing specifically on preservice teacher education. I argue that "the outcomes question in teacher education" (Cochran-Smith, 2000, a, b; in press) is currently driving the field and to a great extent, determining policy and practice. I begin this article by reviewing the policy context, including those reforms and initiatives that have most influenced how outcomes are being constructed, debated, and enacted in teacher education. Then I identify three major "takes" on teacher education outcomes—outcomes as the long-term or general impacts of teacher education, outcomes as teacher candidates' scores on high stakes teacher tests, and outcomes as the professional performances of teacher candidates, particularly their demonstrated ability to influence student learning. For each of these three constructions of outcomes, I consider underlying assumptions about teaching and learning, evidence and criteria used for evaluation, units of analysis, and consequences for the profession. I conclude by considering in some detail the pitfalls and problems that are implicated in various constructions of teacher education outcomes.

The Questions That Drive Reform in Teacher Education

The recent history of teacher education—roughly the last half century—has been analyzed in terms of philosophical and epistemological positions, historical trends, and paradigms of inquiry (Borrowman, 1956; Floden & Buchman, 1990; Griffin, 1999; Klausmeier, 1990; Lucas, 1999; Shulman, 1986; Urban, 1990; Yarger & Smith, 1990; Zeichner, 1988). Another way to think about and trace teacher education reform, however, is in terms of the major questions that have driven the field and the varying and sometimes competing ways these questions are constructed, debated, and enacted in research, policy, and practice.

Along these lines, a very loosely chronological (and necessarily simplified) list of the major questions that have driven teacher education reform over the last 50 years might go something like this: the attributes question, the effectiveness question, the knowledge question, and what I am proposing we now think of as "the outcomes question" in teacher education. Each of these questions both shaped and was shaped by the political climate, the degree and kind of public attention to K-12 schooling, the perceived supply and demand of teachers, federal and state policies and funding programs, perceptions of teacher education as a profession and an area of scholarship that ought to be located (or not) in colleges and universities, and emerging and competing paradigms and programs of research on teaching, teacher learning, and teaching/learning/curriculum in the subject areas.

The Attributes Question

The attributes question, which was prominent from roughly the early 1950s through the1960s, asked, "What are the attributes and qualities of good teachers, prospective teachers, and teacher education programs?" Explored through studies of the personal characteristics of teachers and teacher educators, versions of this question emphasized both attributes related to personal integrity and human sensitivity (the "character" of the teacher or prospective teacher) as well as attributes of the liberally educated and/or academically able person (the "quality" of the teacher or prospective teacher). A different version of the attributes question was central to critiques of teacher education programs and faculty, especially the degree to which they provided (or, more often, failed to provide) intellectually rigorous, discipline-based training for new and experienced teachers worthy of a place in the university. This version of the attributes question animated program decisions and policy debates about the balance between professional versus arts and sciences courses for prospective teachers, the academic qualifications and scholarship (or lack thereof) of teacher education students and faculty, and the organizational structures of teacher education programs.

The Effectiveness Question

The effectiveness question focused different issues: "What are the teaching strategies and processes used by effective teachers, and, what teacher education processes are most effective in ensuring that prospective teachers learn these strategies?" This question drove many of the developments and reforms in teacher education during the late 1960s through the mid 1980s. Influenced by new studies of the "scientific basis of teaching" and by empirical evidence about effective teaching strategies, many teacher education programs developed systems for evaluating prospective teachers according to scientific objectives and stated performance criteria (Gage, 1972). Checklists and other forms of assessment attempted to align classroom teachers' practices with the criteria used by fieldwork supervisors to evaluate the practice of teacher candidates and also with teacher education processes, programs, and language. Some of the other questions that shaped this period arose at least partly in response to perceived flaws in the effectiveness question (Shulman, 1986). New questions rooted in anthropological and sociolinguistic theories about the meanings of classroom events for participants, for example, countered the effectiveness question and pointed to what was left out of discussions that focused on effective teacher behaviors (Erickson, 1986).

The Knowledge Question

Prompted by but also concurrent with public concern about the quality of teaching and teacher education, the knowledge question drove the field from the early 1980s through the late 1990s. This question became mantra throughout the field, "What should teachers know and be able to do?" and/or, its companion, "What should the knowledge base of teacher education be?" At the heart of the knowledge question was the desire to professionalize teaching and teacher education by building a common knowledge base for the profession. Building on early research about teachers' thinking and on emerging knowledge in the various subject matter disciplines related to children's learning, the knowledge question moved the field away from an emphasis on what effective teachers do to a focus on what they know and need to know, the knowledge sources they use, how they organize and evaluate knowledge (Barnes, 1989), and how they learn to construct new knowledge that is appropriate for differing local contexts (Cochran-Smith & Lytle, 1993), particularly for increasingly diverse learners (Banks, 1996).

Versions of the knowledge question identified and made distinctions among formal and practical knowledge (Fenstermacher, 1994), pedagogical content knowledge (L. Shulman, 1987), case knowledge (J. Shulman, 1992), craft knowledge (Grimmett & MacKinnon, 1992); knowledge in action (Schon, 1983), reflection on knowledge (Schon, 1987; Zeichner & Liston, 1987), culturally relevant knowledge (Ladson Billings, 1995; Irvine, 1990), and local knowledge generated through teacher research (Cochran-Smith & Lytle, 1993) and/or action research (Noffke, 1997). Prompted in part by new programs of research and in part by changing accreditation standards, the knowledge question drove major policies and program revisions in teacher education intended to ensure that the burgeoning codified knowledge base was at the center of the curriculum (Reynolds, 1989; Murray, 1996). Some versions of the knowledge question concentrated on the contexts within which prospective teachers could gain the knowledge and practices they need. This question prompted the development of new teacher education contexts, including school-university partnerships (Sirotnik & Goodlad, 1988; Jacobson, et. al, 1998), professional development schools (Holmes Group, 1996; Levine & Trachtman, 1997), and new forms of collaboration among beginning and experienced teachers, teacher educators, and arts and sciences faculty (Goodlad, 1994; Patterson, Michelli, & Pacheco, 1999).

Questioning the Questions

As we close the twentieth century and open the twenty-first, the major question that is driving the field is the outcomes question in teacher education, which I explore in the remainder of this article. Before turning to the outcomes question, however, several other comments are important. First it is important to point out that the questions I have sketched above are not simply research questions, although each of them has research aspects, and several have spawned major programs of empirical study. Each of them also has to do with policy and practice in teacher education and with the intersections as well as disconnects among the three. More important to note, however, is the fact that each of these animating questions is also in some fundamental way a question about the priorities and goals of the profession (and even of the nation). As James Hiebert (1999) points out in a thoughtful article about the relationships between mathematics research and National Council of Teachers of Mathematics (NCTM) standards, the rightness or legitimacy of priorities and goals are questions of value and belief rather than questions of evidence that can suggest educational policies based on varying levels of confidence. Values questions, of course, cannot be settled empirically. It is important to acknowledge, however, that in some cases, policies and practices are driven more by values than by empirical evidence, and, as I indicate throughout this article, all policies and programs of research are ideological in a certain sense.

Second, I want to make it clear that the short list I have offered here does not presume to include the only questions that have driven the field of teacher education nor even necessarily what some people would consider to be the most important questions. There has not been complete consensus in teacher education at any point over the last half century—nor is there now—about which questions are the right ones to ask. There have always been—and hopefully will continue to be—competing questions as well as questions that critique, play off of, and take on the major animating issues. Thus my short list knowingly leaves out a host of important issues and critical questions that have been explored energetically by practitioners, policy makers, and researchers in teacher education.

Finally it is important to note that none of the questions I have loosely associated with particular time periods was settled during that time period or disappeared from consideration after that time. Rather many of the questions that drive the field during particular eras are periodically recycled, reemphasized, and rethreaded into new and current intersections of research, practice, and policy in ways that may or may not appear to be different from their previous iterations. For example, some of the questions about intellectual rigor in teacher education programs and the questionable scholarship of teacher education faculty that were prominent in the late 1950s and early 1960s reemerged in the 1980s (Earley, 2000). Even though the "new" critiques apparently had little to offer that was different from the old (Zeichner, 1988), they *were* nonetheless *different* in that they emerged in the context of a different social and political climate. Similarly, as I suggest below, some of the underlying assumptions of 1970s and 80s questions about the relationships of teaching and learning processes and products (Dunkin & Biddle, 1974) are being recycled into some current versions of the outcomes question in teacher education, and of course some outcomes questions were also explored in the early and mid 1980s. Old questions, however, are never just "same ole" old questions. They are instead "new" old questions because they have a different import and a different set of implications when they are woven into the tapestry of a changed and changing political, social, and economic time.

The Outcomes Question

As we enter the twenty-first century, the outcomes, consequences, and results of teacher education have become critical topics in nearly all of the state and national policy debates about teacher preparation and licensure as well as in the development of many of the privately and publicly funded research agendas related to teacher and student learning. If the major question that drove the field during the last fifteen years was, "What should teachers and teacher candidates know and be able to do?" then the driving question for the last three or four has been, "How will we know when (and if) teachers and teacher candidates know and can do what they ought to know and be able to do?" In the remainder of this article, I elaborate and analyze how policy makers, practitioners, and researchers are constructing the outcomes question in teacher education, examining what I argue are its three major forms. First, however, I briefly consider the larger policy and professional contexts out of which the outcomes question in teacher education emerged and continues to evolve.

Policy and Professional Contexts Of the Outcomes Debate in Teacher Education

The context of reform in teacher education has been analyzed and described at great length from policy (Darling-Hammond, Wise & Klein, 1999; Kaplan & Edelfelt, 1996), curricular (Darling-Hammond & Sykes, 1999; Griffin, 1999), organizational (Jacobson, Emihovich, Helfrich, Petrie, & Stevenson, 1998; Patterson, Michelli, & Pacheco, 1999), and political (Gallagher & Bailey, 2000; Hudson & Lambert, 1997) perspectives. In the section that follows, I sketch the outlines of what might be thought of as the policy and professional context of the outcomes debate in teacher education, or, those reforms and developments in teacher education that have had a strong influence on how the outcomes question is currently being constructed, critiqued, and enacted.

Professionalization of Teaching

First and perhaps foremost, the outcomes debate is deeply embedded in the movement to professionalize teaching and to secure for teaching and teacher education a legitimate place among other health and human services professions. As is now well-documented, there has been a major effort over the last 15 years to codify and disseminate the formal knowledge base for teaching and teacher education in order to insure that teacher education is no longer a normative, natural, or intuitive process (Gardner, 1989). Prompted in large part by nationwide criticisms of teaching and teacher education in the early and mid 1980s (Carnegie Task Force on the Teaching Profession, 1986; Holmes Group, 1986; National Commission on Excellence in Education, 1983) and by early work about teachers' thinking (Clark & Peterson, 1986) and knowledge (Shulman, 1986, 1987), the professionalization movement was intended to make teacher education a state-of-the-art field by establishing an official and formal body of knowledge that distinguished professional educators from lay persons (Gardner, 1989; Yinger, 1999).

The development of standards for the profession has been a central part of the professionalization movement. Since the mid 1980s, the National Council for the Accreditation of Teacher Education (NCATE) has evaluated teacher preparation programs according to the professional knowledge bases and later the conceptual frameworks that shaped and connected the various coursework and fieldwork pieces of the curriculum. The National Board for Professional Teaching Standards (NBPTS) was established in 1987 as the first professional organization in the teaching profession to establish standards for the advanced certification of highly experienced and successful teachers. These were parallel to the model performance-based licensing standards developed by the Interstate New Teacher Assessment and Support Consortium (INTASC), which was initiated in 1987 by the

Council of Chief State School Officers to support the work of states in rethinking and reinventing teacher preparation and teacher licensing (Yinger & Hendricks-Lee, 2000). NCATE 2000 standards also offer performance standards in keeping with those of NBPTS and INTASC (Darling-Hammond, Wise, & Klein, 1999). This means that there are major efforts now well underway to develop a common national system of accreditation of "professionally grounded and performance-based standards for education, licensing, and certification" (Darling-Hammond, Wise, & Klein, 1999, p. 11) that is remarkably broad-based in its support and connects the accreditation of teacher preparation institutions with initial state licensing systems as well as systems for the advanced certification of experienced teachers. All of these center on authentic assessment of teacher performance.

As Yinger argues quite persuasively (Yinger, 1999; Yinger & Hendricks-Lee, 2000), standards always play a critical role in the process of professionalization by establishing public definitions of effectiveness, performance criteria for thinking and action, and goals for initial and continuing professional learning. Notwithstanding the critique that professional standards for teaching and teacher education are largely provisional and unvalidated—based on a consensus of professional educators and an emerging knowledge base rather than on tested outcomes and solid evidence (Murray, 1996, 2000), standards are now part of state licensing requirements in most states are play a major role in the outcomes context.

New Understandings of Teacher Learning

Part of the professionalization of teaching and teacher education was mounting recognition that training models were inadequate to the major tasks of teaching and school reform, and new models of professional development for prospective and experienced teachers were required (Cochran-Smith & Lytle, 1993; Little, 1993; McLaughlin, 1994; Darling-Hammond & McLaughlin, 1995). In fact, as we enter the new century, it is now being suggested that there is a "new paradigm" for professional development and a "new professional consensus" about what teacher education and teacher learning need to look like in order to handle the new tasks of teaching and learning in restructured schools (Darling-Hammond & Sykes, 1999; Hawley & Valli, 1999; Stein, Smith & Silver, 1999). As I have suggested elsewhere (Cochran-Smith & Lytle, 2000), the general orientation of the "new" approach to professional development is more constructivist than transmission- oriented; it is based on the recognition that both prospective and experienced teachers (like all learners) bring prior knowledge and experience to all new learning situations, which are social and specific. In addition, it is now generally understood that teacher learning takes place over time rather than in isolated moments in time, and that active learning requires opportunities to link previous knowledge with new understandings. It also has been widely acknowledged that professional development needs to be linked to educational reform (Loucks-Horsley, 1995) and needs to focus on "culture-building" not skills training (Lieberman & Miller, 1994). It is generally agreed that professional development that is linked to student learning and curricular reform should be embedded in the daily life of schools (Darling-Hammond, 1998; Elmore & Burney, 1997) and should feature opportunities for teachers to inquire systematically about how teaching practice constructs different kinds of learning opportunities for students (Little, 1993; Ball & Cohen, 1999; Cochran-Smith & Lytle, 1993). These new understandings about teacher learning are consistent and intertwined with the emerging standards for the profession noted above.

Standards for Curriculum and Subject Matter Teaching

At the same time that researchers and practitioners in teaching and teacher education were working to build and codify a knowledge base, new frameworks for teaching, learning, and curriculum in almost every K-12 subject area were also being developed by the discipline-based professional organizations such as the National Council of Teachers of Mathematics (NCTM) and the National Council of Teachers of English (NCTE). These were based on new understandings about learning, cognition, and the socio-psycho-cultural construction of subject matter understandings. These were intended to promote teaching for meaning and understanding and explicitly to avoid narrow emphases on skills development and rote learning. New curriculum frameworks were eventually implemented in almost every state, and in most of these, they were coupled with new standards for K-12 student achievement. In most states, new teaching and learning standards were eventually accompanied by high stakes paper and pencil assessments intended to be tightly aligned with the knowledge and skills outlined in the new curriculum frameworks, which in turn were to be tightly aligned with the new knowledge bases in each of the disciplinary areas as established by the professional organizations. Taken together, these developments formed the backbone of the standards movement and what Robert Roth (1996) has called "the age of standards."

National Commission on Teaching and America's Future

Undoubte dly one of the most influential factors in the policy context was the publication in 1996 of *What Matters Most: Teaching for America's Future* (Report of the National Commission on Teaching and America's Future) and the materials that followed it—*Doing What Matters Most: Investing in Quality Teaching* (National Commission on Teaching and America's Future, 1997), *Studies of Excellence in Teacher Education* (Darling-Hammond, 2000, b), and *Promising Practices: New Ways to Improve Teacher Quality* (U.S. Department of Education, 1998). As Gallagher and Bailey (2000) point out, privately commissioned blue ribbon reports such as National Commission on

Teaching and America's Future (NCTAF)—and before it the Flexner Report on medical education and The Reed Report on legal education—have been used since the early part of the twentieth century to call public attention to perceived crises of national importance and to shape the discourse among practitioners, policy makers and the general public. NCTAF's Executive Director, Linda Darling-Hammond, along with colleagues and collaborators in the policy, research, and practice of teacher education, have been explicit and tireless in getting the word out about the central message of the report: what teachers know and can do is the single most important influence on how and what students learn (NCTAF, 1996; Darling-Hammond, 1998 a,b, 2000b; Darling-Hammond, Wise & Klein, 1999; Darling-Hammond & Sykes, 1999; Gallagher & Bailey, 2000). Based on this premise, the policies called for by NCTAF, many of which are now being implemented in states across the country, is exquisitely clear:

> We propose an audacious goal for America's future. Within a decade—by the year 2006—we will provide every student in America with what should be his or her educational birthright: access to competent, caring, qualified teaching in schools organized for success (NCTAF, 1996, p. vi).

NCTAF's now highly familiar list of recommendations includes: getting serious about standards for students and teachers; reinventing teacher education and professional development; placing qualified teachers in every classroom in America; supporting and rewarding teachers' developing knowledge and skill; and creating schools organized to support and sustain student and teacher success.
What is unprecedented about the commission's report is the call for all of its recommendations to be addressed in concert in order to achieve across the states a coherent and consistent system of reform in teacher education, teacher licensing, and teacher accreditation (NCTAF, 1997). This requires consistency across several major efforts, including the move toward performance-based standards for teacher licensing, parallel efforts to develop authentic assessments of teachers, and the development of national standards for teacher education, licensing, and certification. These national efforts are being led by NCTAF, NBPTS, INTASC, and NCATE (Darling-Hammond, Wise, & Klein, 1999).

Also unprecedented are the teeth that the NCTAF recommendations now have in terms of federal money and policy related to professional development, teacher education, and federal grants (Earley, 2000). In 1997, the Department of Education sponsored a five year, $23 million consortium of research universities and professional organizations in order to develop a research base supporting the implementation of recommendations put forth by NCTAF. In 1998 the Higher Education Act (HEA) was signed into law; of particular importance in terms of the policy context for the outcomes debate are the mandatory (but unfunded) accountability requirements for states and higher education institutions contained in Title II (Earley, 2000). These require that all states and colleges/universities that receive any federal dollars through HEA must provide annual information on the performance of all teacher candidates recommended by an institution on each measure required for licensure. These data will be compiled into institutional and state report cards intended to serve as indicators of "the health of the teacher education enterprise" (Earley, 2000), which will provide public rankings of each teacher education institution .

New Standards for Teacher Education Accreditation

What is closest to day-to-day work of teacher educators are the new outcomes-based approaches to evaluating teacher preparation programs and institutions. An outcomes-based approach is now in effect at NCATE (1999), the major teacher education accrediting agency. Emphasizing outcomes rather than inputs was also a major reason for the founding of newcomer accrediting organization, Teacher Education Accreditation Council (TEAC) (Teacher Education Accreditation Council, 1999). Although fewer than half of the nation's teacher preparation institutions are currently accredited, NCATE-accredited institutions produce two thirds of the nation's teachers. In addition, NCATE has relationships with 40-some states, and some are moving to require all teacher preparation institutions to seek accreditation from either NCATE or TEAC (Wise, 1999).

In recent articles and symposia, NCATE 2000's new focus on outcomes has been described as a "paradigm shift from inputs to outputs" (AACTE, 2000), a "bold" and "daring… plunge into the world of performance assessment and performance standards" (Schlalock & Imig, 2000, p. 4), and a "major shift from curriculum- oriented standards to performance-based standards that focus on what teacher candidates know and are able to do" (Wise, 1999, p. 5). NCATE's prior standards were described by critics as merely "counting courses" or focusing on curriculum content instead of paying attention to results. The new standards focus on what teacher candidates can actually do in schools and classrooms by emphasizing performance, particularly in relation to students' learning. The new standards, which received final approval in 2000, are effective for all institutions seeking NCATE accreditation during or after Fall 2001. NCATE's new system will require schools of education to provide performance evidence of candidate competence, including state licensing examination results as well as summarized and sampled performance evidence of candidates' knowledge and skill (Wise, 1999). The stated rationale for the first major section of the new standards, "Candidate Performance," makes this emphasis clear:

> The public expects that teachers of their children have sufficient knowledge of content to help all students meet

> standards for P-12 education. The teaching profession itself believes that student learning is the goal of teaching. NCATE's Standard 1 reinforces the importance of this goal by requiring that teacher candidates know their content or subject matter, can teach, and can help all students learn . . . Candidates for all professional education roles are expected to demonstrate positive effects on student learning. Teachers and teacher candidates should have student learning as the focus of their work...Primary documentation for this standard will be candidates' performance data prepared for national and/or state review ...[including] performance assessment data collected internally by the unit and external data such as results on state licensing tests and other assessments. (NCATE, 1999, pp. 7-9)

The new NCATE standards are in keeping with movement to professionalize teaching and also consistent with recent developments in specialized accreditation organizations more generally, where the emphasis has shifted from inputs to outcomes measures (Dill, 1998). This is part of a larger trend in higher education, what Graham, Lyman and Trow (1995) refer to as an "increasing clamor to apply quantitative measures of academic outcomes to guarantee educational quality for consumers" (p. 7) at the higher education level.

The Deregulation Movement

The aspects of the policy context for the outcomes debate that I have mentioned so far are in sync with one another in certain important ways —the development of standards for subject matter teaching, new understandings of teacher learning, new standards for the accreditation of teacher education institutions, and the efforts of NCTAF, NBPTS, INTASC, and NCATE to unify teacher preparation, licensing, and certification. All of these are consistent with the first item on the list—the movement to establish teaching (and teacher education) as a legitimate profession with a well-established knowledge base (Reynolds, 1989; Murray, 1996; Houston, 1990; Sikula, 1996), jurisdictional responsibility for defining and acting on professional problems (Yinger, 1999; Yinger & Hendricks-Lee, 2000), and clear principles or standards for professional practice (NCTAF, 1996; Darling-Hammond, Wise & Klein, 1999). Each of these initiatives works from but also builds on the dual premises that caring, competent, and qualified teachers are essential to insuring rigorous learning opportunities for all children in America's schools *and* that upgrading teacher education and credentialing for the profession are necessary for ensuring that all children have such teachers.

As is now well known, however, the professionalization movement is not the only national agenda related to teaching and teacher education. There is also a well publicized and well-funded movement to deregulate teacher education by dismantling teacher education institutions and breaking up the monopoly that the profession (i.e., schools of education, professional accrediting agencies, and many state licensing departments) has, according to its critics, too long enjoyed. The deregulation movement, well-funded by conservative political groups like the Heritage Foundation, the Pioneer Institute, and the Fordham Foundation, begins with a premise that is radically different from the premises of professionalization. Those who support deregulation assert that teacher education programs and most of the requirements of state licensing agencies are unnecessary hurdles that keep bright young people out of teaching and focus on social goals (even "social engineering") rather than academic achievement (Kanstoroom & Finn, 1999).

Denigrating professionalization efforts as the "romance of regulation" (p. 3), the Fordham Foundation's 250 page volume on how to get "better schools" and "better teachers" (Kanstoroom & Finn, 1999), for example, intentionally frames its agenda in opposition to efforts to professionalize teaching and teacher education. The Fordham Foundation "manifesto" asserts:

> Today in response to widening concern about teacher quality, most states are tightening the regulatory vise, making it harder to enter teaching by piling on new requirements for certification. On the advice of some highly visible education groups, such as the National Commission on Teaching and America's Future, these states are also attempting to 'professionalize' teacher preparation by raising admissions criteria for training programs and ensuring that these programs are all accredited by the National Council for the Accreditation of Teacher Education (NCATE). That organization is currently toughening its own standards to make accredited programs longer, more demanding, and more focused on avant-garde education ideas and social and political concerns…
>
> The regulatory strategy that states have followed for at least the past generation has failed. The unfortunate results are obvious: able liberal arts graduates avoid teaching, those who endure the process of acquiring pedagogical degrees refer to them as 'Mickey Mouse' programs, and over time the problems of supply and quality have been exacerbated. When a strategy fails, it does not make much sense to do the same thing with redoubled effort. Yet that is what many states are now doing. (pp. 4- 5)

Lest anyone think they eschew all regulations related to teacher education, editors of the Fordham volume concede that some regulation is necessary:

> Every child should be able to count on having a teacher who has a solid general education, who possesses deep subject area knowledge, and who has no record of misbehavior. The state has an obligation to ensure that all prospective teachers meet this minimal standard. (p. 11)

Publications by Chester Finn and colleagues (e.g., Kanstoroom & Finn, 1999; Finn, Kanstoroom, & Petrilli, 1999; Klagholz, 2000; Finn & Petrilli, 2000) advocate alternate routes into teaching, high stakes testing as the primary way to ensure teachers' subject matter knowledge, and a heavy emphasis in schools on academic achievement, order, and discipline (Farkas & Johnson, 1997). Part of a larger conservative political agenda for the privatization of American education, the deregulation movement is an influential part of the policy context in teacher education and, as I argue here, it is playing a major role in the ways we construct outcomes in teacher education.

Sorting Out the Outcomes Question

The different ways outcomes are being constructed in teacher education rest on differing assumptions about what teachers and teacher candidates should know and be able to do, what K-12 students should know and be able to do, what counts as evidence of "knowing" and "doing," and what the ultimate purposes of schooling should be. Different premises about the purposes of schooling mean different ways of demonstrating that teacher education programs and procedures are "accountable," "effective," or "value-added." Despite these differences, however, most discussions about teacher education outcomes have to do with the connection between teacher education and student learning. In a certain sense, every debate related to outcomes assumes that the ultimate goal of teacher education is student learning and that there are certain measures that can be used to indicate the degree to which this outcome is or is not being achieved by teacher candidates, K-12 students, teacher educators, higher education institutions, local or state policies, and the education profession itself. At a general level, then, the outcomes debate in teacher education revolves around these two questions:

> What should the outcomes of teacher education be for teacher learning, professional practice, and student learning?
>
> How, by whom, and for what purposes should these outcomes be documented, demonstrated, and/or measured?

It is important to note that unanimity about the outcomes questions we should be asking begins and ends here, at this rather surface level of understanding. If we move one level deeper in terms of specificity or elaboration, we uncover disagreement. If we attempt to describe the relationship between teacher learning and professional practice, attempt to explain what we mean by teacher learning and student learning, attempt to elaborate the theoretical bases and consequences of the kinds of student learning we are trying to account for, or even attempt to define what we mean by "students" (which students? how many? all of them or some statistically significant portion of them?), we uncover differences, some of which represent deep philosophical and political divides. Notwithstanding the growing—and many say unprecedented—consensus about standards for teaching and teacher education (Darling-Hammond, 1996, 2000; Darling-Hammond, Wise & Klein, 1999), it is important to acknowledge that there is considerable variation both within and outside the profession in terms of how outcomes are being constructed and upon what grounds they are being debated.

The question of outcomes is being taken up in differing ways depending on the policy, research, and practice contexts in which it is posed as well as on the political and professional purposes of the posers. One way to sort out different ways of constructing teacher education outcomes is to consider at least the following:

1. How are "teacher learning," "professional practice," and "student learning" defined, or, what is used as a proxy for these? How are teacher learning, professional practice, and student learning assumed to be related to one another? What is assumed to be central or extraneous?
2. What counts as evidence of teacher learning and student learning? What are the criteria against which the evidence is measured? What is the source of these criteria? What is the unit of analysis?
3. What is assumed to be the larger purpose of schooling and the role of schooling in society?
4. What is the larger political and/or professional agenda behind a given construction of outcomes? What are the consequences for policy and practice of constructing outcomes this way?

As Figure 1 indicates, at least three major ways of constructing outcomes in teacher education are currently receiving major attention and visibility nationally, at the state level, and within teacher education institutions: the long-term or general impacts of teacher education as a profession; the aggregated scores on teacher tests of teacher candidates, teacher preparation programs, and/or higher education institutions; and the professional performances expected of teachers and teacher candidates. In some policy and practice contexts, one or more of these is used in combination with others to guide decisions about distribution of resources, licensing and accreditation privileges, and relative rankings of programs, institutions, and individuals.

Figure 1

Constructing Outcomes in Teacher Education: Three "Takes" on the Outcomes Question

The Outcomes Question in Teacher Education	What should the outcomes of preservice teacher education be for teacher learning, professional practice, and student learning? How, by whom, and for what purposes should these outcomes be documented, demonstrated, and/or measured?
Outcome as "long-term/general impact"	What long-term and/or general impacts should preservice teacher education be expected to have, particularly on student achievement?
Outcome as "teacher test results"	What impact should preservice teacher education be expected to have on teacher test results? What results on teacher tests should be expected of teacher candidates, teacher education programs, higher education institutions, states?
Outcome as "professional performance"	What professional performances should teacher candidates be expected to demonstrate? How should teacher candidates and teacher education programs/institutions be expected to document, analyze, and evaluate these professional performances?

So far in this article, I have explained *why* the outcomes question is the question that is driving reform in teacher education at this particular juncture of political, professional, and social contexts. In the next section, I take each of the major "takes" on the outcomes questions and look more closely at *how* they are being constructed in teacher education and then consider *what the consequences (and pitfalls)* of these constructions are for policy and practice.

Long-term/General Impact as Outcome of Teacher Education

The first major take on the outcomes question concerns the long-term or general impact of teacher education on teacher knowledge, teacher preparedness, teacher attrition, teacher ratings, and student achievement. Explorations of these questions in teacher education are located within much larger debates about teacher quality and teacher qualifications, teacher licensing and certification, professional standards for teaching and curriculum, and the use of student achievement as a valid evaluation measure for teachers and schools. Various studies have analyzed whether teacher candidates who have completed approved teacher education programs stay in teaching longer than those without such preparation, whether their attitudes and knowledge about teaching and learning are different (Ashton & Crocker, 1987), whether they feel more committed to teaching than others or more prepared to teach, and whether their principals rate them higher or lower than others (Haberman, 1985). Studies have also compared the teaching ratings of liberal arts graduates with those prepared in pedagogy (Haberman, 1985; Grossman, 1990) and/or have compared the teaching effectiveness, including the classroom management skills, of those with minimal versus extensive subject matter knowledge and/or minimal versus full preparation in teaching (Ashton & Crocker, 1987; Evertson, Hawley & Zlotnik, 1985; Kennedy, 1991; Denton & Lacina, 1984; Darling-Hammond, 1991). Other studies have considered whether education and subject matter preparation predict "teaching performance" of teacher candidates (Ferguson & Womack, 1993) and/or have an impact on students' achievement (Ashton & Crocker, 1987). There is a great deal of attention currently to sorting out the results of these studies and drawing policy conclusions from them.

As we enter the new century, the issue that is most visible *and* most highly contested has to do with the impact of teacher education on K-12 students' learning. This question, debated in the research literature and in the media, is being explored primarily through meta-analyses and/or syntheses of previous and current work in order to make recommendations about teacher education as state policy that is either value-added or not, either a good investment or not. In these high stakes debates, teacher education at the preservice level is not considered by itself but as one of several factors related to the quality and qualifications of teachers. The unit of analysis is not teacher candidates—individually or collectively—or even teacher preparation programs and institutions. Rather the unit of analysis is the profession itself—teacher preparation as one aspect of a broad category referred to as "teacher qualifications," which includes scores on licensure examinations, graduate level degrees, years of experience, preparation in the subject matter area of certification as well as in pedagogy, type and extent of certification in the teaching area, and amount of money spent by school districts on professional development. Student learning is generally defined as student gains on achievement tests, often reading and mathematics in grades one through twelve. The relationship between the two is taken to be the percentage of variance in student gains accounted for by teacher qualifications when other variables are held constant or adjusted. The pertinent units of analysis are aggregated student achievement scores and general indices of teacher qualifications that include multiple features.

Questions about the long term impact of teacher education are at the heart of many policy debates related to the initial preparation of teachers as well as teachers' continuing professional development. These have enormous implications for how states (and now the federal government) support and invest in the improvement of schooling, how higher education institutions support and invest in teacher education programs and schools of education, and how school districts establish and maintain hiring and reward systems as well as local programs of ongoing professional development.

Synthesizing the Research: "Teacher Education Matters Most"

The initial report of NCTAF (1996) addressed the question of long-term impact directly by linking teacher qualifications—including extent of teacher education— with student learning. Speaking for the Commission, Darling-Hammond (1998) argued that a growing body of research "appears to confirm" that teacher knowledge and teacher expertise are significant influences on student learning, as are to a lesser extent class size and school size. Although Darling-Hammond pointed out that the initial Commission Report was a starting point for more public discourse rather than a set of research-based conclusions, this work was widely cited by those committed to elevating the status of the teaching profession, particularly by those embroiled in battles about teacher certification regulations at the state level.

The NCTAF report was highly successful in generating public discourse about teaching and teachers—Darling-Hammond (2000) indicates that more than 1500 news articles and editorials have appeared nationally and internationally since its publication. Major research syntheses that support the initial direction of the report (Darling-Hammond, 1998, 1999, 2000b; Sykes & Darling-Hammond, 1999) have also now appeared as have several case studies (e.g. Elmore & Burney, 1997) that provide contextual information. Darling-Hammond's (2000, b) major synthesis of research on teacher quality and student achievement has been disseminated widely. The synthesis, which appeared in this electronic journal on January 1, 2000, had been retrieved more than 23,000 times year later. This review provides what is probably the clearest example of how long-term impact is being constructed as an outcome of teacher education; the review explores the impact on students' achievement of large scale policies and institutional practices that affect the overall level of teachers' knowledge and skills in a given state or region.

Drawing on data from an NCTAF 50-state survey of policies, case studies at the state level, the 1993-94 Schools and Staffing Surveys (SASS), and the National Assessment of Education Progress (NAEP), Darling-Hammond (2000b) examined how teacher qualifications are related to students' achievement. She concluded:

> The findings of both the qualitative and quantitative analyses suggest that policy investments in the quality of teachers may be related to improvements in student performance. Quantitative analyses indicate that measures of teacher preparation and certification are by far the strongest correlates of student achievement in reading and mathematics, both before and after controlling for student poverty and language status. . . This analysis suggests that policies adopted by states regarding teacher education,

> licensing, hiring, and professional development may make an important difference in the qualifications and capacities that teachers bring to their work. (p. 1)

Constructing the outcomes of teacher education as long-term impact on students' achievement is part of NCTAF's larger campaign to provide qualified and competent teachers for all students by emphasizing and aligning professional standards across initial teacher preparation, teacher licensure, and teacher certification at the state and regional levels. This take on the outcomes question provides little information about the impact of teacher education disaggregated from teacher qualifications more generally, nor does it address the relative merit of various approaches to teacher education, although there is related research that does so. But this was never the point of constructing outcomes as long-term impact of teacher qualifications on students' achievement. The point was to demonstrate that teacher education, as part of teacher professionalization more broadly, was and is a good investment—for state policy makers, for higher education institutions, and for the future of a democratic society.

Synthesizing the Research: "Teacher Education Doesn't Matter Much"

There is, however, another conclusion about long-term impact as an outcome of teacher education. Economists such as Dale Ballou, Michael Podgursky, and others (Ballous & Podgursky, 1997, 1998, 1999; Goldhaber & Brewer, 1999) offer analyses of teacher preparation, licensing and certification that support the deregulation of teacher education and seek to limit the power of the educational community to control the profession. For example, in what they refer to as a "layman's guide" to teacher training and licensure that appears in the Fordham Foundation's (Kanstaroom & Finn, 1999) policy statement on how to produce better teachers and better schools, Ballou and Podgursky (1999) conclude:

> [T]eacher ability appears to be much more a function of innate talents than the quality of education courses. Teachers themselves tell us that this is so. We come to similar conclusions when we examine the determinants of scores on teacher licensing examinations. Finally, teachers who enter through alternative certification programs seem to be at least as effective as those who completed traditional training, suggesting that training does not contribute very much to teaching performance, at least by comparison with other factors. (p. 57)

Like the syntheses that support the recommendations of NCTAF, the summaries by these conservative economists construct outcomes in teacher education as part of a general category of teacher qualifications (including teacher preparation and licensing based on completion of accredited programs) and in terms of student achievement and teacher attrition. They draw in many instances on the same data and even refer to many of the same sources that are used by Darling-Hammond and others.

Despite a certain surface level of similarity, however, the deregulation-ists reach conclusions that are diametrically opposed to the conclusions of those who advocate professionalization. The introduction to the Fordham Foundation's policy statement (Fordham Foundation, 1999), which is signed by William Bennett, Chester Finn, E.D Hirsch, James Peyser, and Diane Ravitch, among others, states this conclusion in no uncertain terms:

> We are struck by the paucity of evidence linking inputs [courses taken, requirements met, time spent, and activities engaged in] with actual teacher effectiveness. In a meta- analysis of close to four hundred studies of the effect of various school resources on pupil achievement, very little connection was found between the degrees teachers had earned or the experience they possessed and how much their students learned. (p. 18)

Contrast this conclusion with Linda Darling-Hammond's conclusion in *Doing What Matters Most: Investing in Quality Teaching* (1997):

> Reviews of more than two hundred studies contradict the long-standing myths that 'anyone can teach' and that 'teachers are born and not made' . . .teachers who are fully prepared and certified in both their discipline and in education are more highly rated and are more successful with the students than are teachers without preparation, and those with greater training…are more effective than those with less. (p. 10)

The fact that some of the same evidence is used to make two exceedingly different cases about teacher education is confusing to say the least. (Note 1) Debates about the evidence concerning the relationship of teacher education and student learning outcomes continue, and they are growing increasingly heated. In a recent issue of *Teachers College Record*, for example, Ballou and Podgursky (2000) directly attacked the Commission's findings, and Darling- Hammond (2000) emphatically refuted their use of evidence and their conclusions. Questions about the evidence were also explored in a face-to-face debate between Linda Darling-Hammond and Chester Finn, which was sponsored by the Education Commission of the States (Education Commission of the States, 2000).

The Problem of Teacher Education

Part of the difference in conclusions about the long- term outcomes of teacher education may lie in the details of the ways terms are defined and data are selected for these analyses. For example, there are major differences across reports in what is included under "alternate programs," what it means to be "fully qualified," or "to have a major" in one's area of certification. The accumulation of many small differences in definitions of terms and data analysis procedures may account for some of the major statistical differences and the contradictory conclusions of these two major syntheses. But the differences may also be partly explained by differences in the way "the problem" of teacher education is framed in the first place and how these different constructions shape the ways terms are defined, procedures are established for data selection, results are manipulated, and interpretive frameworks are developed.

Penelope Earley (2000) makes an incisive point along these lines in a recent discussion about the value-laden nature of educational research and its easy use by policy makers to further their own agendas. She suggests that "data and evidence used in the policy process will have several levels of bias: that embedded in the data or evidence itself, bias associated with analysis, and the biases of those in the policy world who use the information" (p. 35). This understanding of the policy process may help to explain some of the differences I have just been highlighting. Supported by the Carnegie Foundation and the Ford Foundation, NCTAF (in collaboration with NBPTS, INTASC, and NCATE) frames "the problem" of American education in terms of democratic values (Engle, 2000; Earley, 2000; Labaree, 1997)) and thus begins—and ends—with calls for stepped-up, standards-driven improvements in teacher education and professional development in order to guarantee a well-qualified teacher for every American school child.

The Fordham Foundation and other conservative organizations and politicians, on the other hand, frame "the problem" in terms of a market approach to educational policy making. They criticize the profession's "preoccupation with teacher preparation" (Ballou & Podgursky, 1997, p.4), seek to limit the power of the profession to control the market by controlling licensing and approved programs, and push an agenda based on what Earley calls "competition, choice, winners and losers, and finding culprits" (Earley, 2000, p. 36). They thus begin—and end—with calls for alternate routes to certification and for eliminating "needless barriers" to the profession. They advocate heavy emphasis on the results of education and favor heavy sanctions for those who cannot or will not measure up. (I return to this issue of market versus democratic ideologies in the final section of this article where I suggest, following many others, that these two approaches to educational policy—democracy-driven and market-driven—are mutually exclusive.)

Teacher Test Scores as Outcome of Teacher Education

The teacher tests now required for initial licensing in most U.S. states (Digest of Educational Statistics, 1997) suggest another highly visible way that outcomes are being constructed in teacher education. The construction of test scores as outcomes is in a certain sense a subset of the preceding construction in that the test scores of prospective teachers are often taken to be one facet of the long-term impact of teacher education. However, because teacher tests have been given so much recent attention and weight, it is worth considering them separately. Debates about teacher tests are connected to larger debates about quality, licensing, standards, and assessment. Teacher tests are also related to the long history of criticisms of teachers as mediocre students, "semi-skilled" workers, "less than literate" individuals, and members of a minor or "not quite" profession.

With initial licensing tests, what is measured (and taken to be an indication of what prospective teachers have learned) is usually some combination of general knowledge, including communication and literacy skills, with knowledge of specific subject matter and pedagogy, both of which are demonstrated on a paper and pencil exam. Although teacher test scores have probably received more publicity and more public outcry than any other recent measure of outcomes, they are linked to teacher performance and K-12 student learning primarily through presumption rather than empirical evidence and/or are considered in combination with other measures of teacher expertise or teacher qualifications that are difficult to untangle as I noted a moment ago. There is little evidence that large-scale implementation of statewide teacher testing programs is affecting the actual classroom performance of teachers (Flippo, 1986; Ladson-Billings, 1998), although there is some evidence that testing has an impact on the "quality" of those entering and remaining in teaching where "quality" is defined as other test scores, grade point averages, and similar measures (Gitomer & Latham, 2000)

Until recently teacher test scores were assumed primarily to measure individual fitness for teaching the way SATs and GREs are assumed to measure individuals' potential for college and graduate level academic work. Relatively little attention was paid to the aggregated scores of individuals from the same state or the same teacher education institution. Times have changed, however, fueled in part by the dismal performance of Massachusetts teacher candidates on that state's first ever teacher test in 1998—when 59% of candidates failed, and Massachusetts House Speaker Thomas Finneran called test takers "idiots" (Melnick & Pullin, 2000). The Massachusetts scores fanned the debate about teacher quality and teacher preparation that was already going on in the U.S. Congress partly in response to the report of NCTAF and in light of proposed stipulations of the reauthorized Higher Education Act. (See Earley, 2000, for an excellent discussion of federal policy debates regarding teacher education and Melnick & Pullin, 2000, for thoughtful analyses of many of the legal and policy issues involved in the Massachusetts teacher test.)

Of particular importance in the Higher Education Act are the mandatory accountability requirements, which stipulate that all states and colleges/universities receiving federal dollars must provide annual information on the performance of all teacher candidates recommended by an institution on each measure required for licensure. As has been widely broadcast, these data are to be compiled into institutional and state "report cards" intended to serve as indicators of the fitness of the teacher education enterprise and will provide public (and no doubt highly politicized) rankings of teacher education institutions in the U.S. (U.S. Department of Education, 2000).

By switching the unit of analysis from individuals to institutions, recent testing arrangements locate the responsibility for teacher education outcomes squarely at the feet of colleges and universities, some of which will be seriously threatened with closure when the new regulations go into effect (Schrag, 1999; Wise, 1988). In some states, it has even been suggested that a major result of teacher tests has been to discredit schools of education and provide ammunition for those who would like to close them (Cochran- Smith & Dudley-Marling, in press). In a strange sort of contradiction, teacher tests in some places are now being framed in the media as both outcomes of teacher education (i.e., teacher education programs and institutions get public blame for low test scores), and, at the same time, prerequisites for teacher education programs (i.e., candidates in some institutions are now being required to take certain portions of tests in order to be admitted to programs in the first place).

Constructing outcomes in teacher education as scores on teacher tests creates a number of problems and has important consequences for the pool of candidates entering the profession. Some statewide teacher tests, for example, are anathematic to the concepts and knowledge taught in teacher education programs (Melnick & Pullin, 2000), particularly in terms of conceptions of literacy, views of student learning, and notions of growth and progress (Luna, Solsken, & Kutz, 2000). Unfortunately, at exactly the same time that we are supposedly interested in recruiting a more diverse pool of teacher candidates, teacher tests are working as gate keepers to keep some potential teachers out. Fear of poor performance on teachers tests is leading some schools of education to change admissions standards with the consequence that fewer students are applying, and there is increasing evidence that the implementation of teacher tests—like other tests historically that are biased against minorities—may be playing a role in the decline of minority participation in the teaching profession (Garcia, 1986; Gitomer & Latham, 2000; Smith, 1984; Wise, 1988). Further, although some studies have also considered whether teacher candidates prepared in fully-accredited teacher education programs (particularly at NCATE-accredited institutions) score higher on teacher tests than those prepared in other teacher education programs and/or those with no teacher preparation (Wise, 1999), there is little evidence that teacher test scores are related to actual teaching performance in classrooms or to students' learning.

Professional Performance as Outcome of Teacher Education

The third take on the outcomes question—and the one that is closest to the everyday work of many teacher educators—has to do with the professional performances that teacher candidates should be expected to demonstrate, including the ways candidates and teacher educators document, analyze, and evaluate these performances. This version of outcomes is located within larger debates about authentic assessments of teaching that result in student learning, the shift from "inputs" to "outputs" as the basis of professional accreditation reviews of teacher education institutions, the development of quality assurance mechanisms based on professional standards that are consistent across the professional lifespan, and a growing body of literature that examines the relationships of inquiry, knowledge, professional practice, and teacher education pedagogy.

Teacher Candidates and Professional Performance

Constructing teacher education outcomes in terms of the professional performances of teacher candidates begins with the premise that there is a professional knowledge base in teaching and teacher education based on general consensus about what it is that teachers and teacher candidates should know and be able to do. The obvious next step, then, is to ask how teacher educators will know when and if individual teacher candidates know and can do what they ought to know and be able to do. A related and larger issue is how evaluators (i.e., higher education institutions themselves, state departments of education, or national accrediting agencies) will know when and if teacher education programs and institutions are preparing teachers who know and can do what they ought to know and be able to do.

In a recent historical sketch of performance assessment, Madaus and O'Dwyer (1999) suggest that today's emphasis on performance assessment in K-12 education is part of a larger sea change in educational measurement that highlights the "3 P's—performance, portfolios, and products" and that has captured "the linguistic high ground, just as the term 'minimum competency testing' did in the 1970s" (p. 688). Madaus and O'Dwyer point out that despite the hype, performance assessment is based on the same technology as all assessments—obtaining a small piece or sample of a candidate's behavior drawn from the larger domain of knowledge and skill it is assumed to be part of and then using the candidate's performance on that sample to make inferences about his or her likely performance on the entire domain. Defining performance assessment broadly, Madaus and O'Dwyer include three ways to sample behavior from a larger domain—requiring an examinee to construct or supply oral or written answers to some set of questions, requiring him or her to perform an act that will be evaluated according to certain criteria, or requiring him or her to produce a product of some kind.

Notwithstanding the long list of cautions about the use of performance assessments for high stakes contexts cited by Madaus and others (Madaus & O'Dwyer,1999; Madaus, 1993; Haertel, 1999), all signs indicate that the teacher education profession is driving full throttle into the world of performance assessment. This is being done for two different purposes, each drawing on different units of analysis: (1) for the purpose of evaluating individual prospective teachers where the unit of analysis is the individual teacher candidate and the evaluator is some combination of school- and university-based teacher educators involved in the candidate's educational program, and (2) for the purpose of evaluating individual teacher education programs where the unit of analysis is the teacher education program itself within and in relation to its larger institutional unit (university, school, college, or department) and where the evaluator is a national accrediting agency, a state department of education, or some combination of the two.

In teacher education, performance assessment is intended to evaluate teacher candidates' ability to produce "products" and complete "authentic tasks" that closely resemble the real work of teaching and do so in ways that are aligned with consistent internal and external standards and criteria. The notion of professional performance as outcome is a central to new partnerships among accrediting, licensing, and certification agencies across states and the nation (Wise, 1996). Performance as outcome is also implicated in the debate between NCATE and TEAC as accrediting agencies, including disagreements about whether the latter is a threat to professionalization or a useful and appropriate accrediting alternative for many institutions (Murray, 2000; Darling- Hammond, 2000, a). Performance as outcome is behind the move in some states to require all teacher education institutions to seek either NCATE or TEAC accreditation as well as other new state requirements that teacher education programs provide evidence that teacher candidates have state-of-the-art knowledge and a demonstrable impact on K-12 students' learning (Wise, 1999).

In the following section I briefly describe four teacher education initiatives or ongoing projects that illuminate how professional performance is being constructed as an outcome of teacher education. Although they use different language, each of these elaborates a process for documenting the linkage between teacher education, teaching practice, and student learning. Each of the programs I use as illustrations here has been highly visible and thus open to public scrutiny as a result of multiple publications and presentations. Each has also been supported by or connected to larger professional foundations, agencies, or organizations and/or has been used as a public exemplar of teacher education practice in keeping with a particular agenda. Taken together, the four examples reveal some of the range and variation in performance as outcome in terms of definitions of teaching and learning, how aspects of teaching are related to one another, and the larger social and political agendas to which teachers' work is attached (or not). Despite differences, however, these examples also reveal some basic similarities in the performances teacher candidates are being required to demonstrate in preservice education. (Note 2)

Ability-Based Performance Assessment

Alverno College's standards-based approach to performance assessment for preservice teachers is part of the larger ability-based curriculum of the college, which was developed in the 1970s in order to meet the needs of a non-traditional student population (U.S. Department of Education, 1998). The work at Alverno College, which specifically eschews curriculum as "counting courses" and fosters instead a view of ongoing "assessment as learning" (Diez & Hass, 1997), has received considerable attention in the literature on outcomes in teacher education (Diez & Hass, 1997; Diez, 1996, 1997, 1998; Alverno College, 1996; Blackwell & Diez, 1999). It has been widely cited and used as an exemplar of preservice teacher education in line with the standards- based professionalization efforts of NCTAF, INTASC, NBPTS, and NCATE (Darling-Hammond, Wise & Klein, 1999; Diez, 1998; National Commission on Teaching and America's Future, 1997). In addition, the U.S. Department of Education's guide to improving teacher quality (U.S. DOE, 1998) features the program at Alverno as one of three preservice programs that exemplify "promising practices," and the *Studies of Excellence in Teacher Education* series co- published by AACTE and NCTAF (Darling-Hammond, 2000, c) include it in their booklet on preparation at the undergraduate level,.

Alverno College's program, which focuses on "what students can do with what they know" (Diez & Hass, 1997, p. 17), is based on the idea that performance assessment is not an add- on, but a basic approach that transforms the curriculum as well as the ways teacher education faculty think about their work. The Alverno curriculum specifies eight general abilities including communication, analysis, problem solving, values within decision making, social interaction, global perspectives, effective citizenship, and aesthetic responsiveness that cut across the entire four year curriculum (U.S. DOE, 1999). Teacher education students also have professional abilities that they must demonstrate including integrating content knowledge with teaching pedagogy, diagnosing individual student needs, and managing resources effectively. Each course has specific goals aligned with general outcomes and requires "complex evidence of student performance."

Students' abilities are assumed to be developmental and, because the evidence they require is complex, assumed to demand multiple opportunities for demonstration of abilities and a wide variety of assessment modes (Diez & Hass, 1997). Thus students are engaged in literally hundreds of performances during their preservice preparation, each of which includes a self-assessment component. In describing the Alverno program in the studies of excellence series, Zeichner (2000) comments, "I doubt that there is a teacher education program anywhere that gives such careful attention to assessment of its students" (p. 11). Performance assessments are "situated in authentic contexts and teaching roles" (Diez & Hass, 1997, p. 21) and based on "proofs" of professional ability such as essays, letters, position papers, case study analyses, observations of teachers, simulations with parents and others, and development of curriculum materials. Program developers point out:

> Alverno faculty believe that performance assessments are most beneficial when they come as close as possible to the realistic experiences of the practicing teacher. In developing the curriculum for teacher education, they have identified a number of roles that teachers play, including but going beyond the primary role of facilitator of learning in the classroom. Therefore, performance assessments of the abilities of a teacher may be simulated to focus on parent- teacher interaction, multidisciplinary team evaluation, the teachers' work with district or building planning, or the teacher's citizenship role, as well as on actual classroom teaching performance in the field experience and student teaching classrooms. In this way they provide candidates with successive approximations of the role of the teacher (Diez & Hass, 1997, p. 24).

The portfolio interview assessment is the major external assessment and is required in order to conclude the pre- professional stage of the program and begin the student teaching period (Zeichner, 2000). Here students compile all of their own work, lesson and unit plans, videotapes of lessons, and self assessments. Portfolios are reviewed by faculty advisors as well as teams of principals and teachers, whose feedback is used to prepare for student teaching.

Chief spokesperson for the program, Mary Diez (2000) emphasizes that Alverno's approach to performance assessment is based on the idea that teaching and learning have to be connected when teaching performance is assessed, especially how particular teaching practices facilitate students' learning and how teachers learn to examine their own and their students' work over time. Like the emphasis of the INTASC and NBPTS standards, the work at Alverno emphasizes how a teacher's thinking leads to improvements in teaching and students' learning. Thus the performances that are required of teacher candidates must indicate teacher learning as much as and in connection to student learning. Through portfolios, analyses of lessons and units, and other self-assessments and reflective activities, teachers learn to look at and make sense of students' work and document the impact of their own practice on students' learning. They are required not simply to demonstrate that their teaching has an impact on students' learning, although they must do that, but also how and why their teaching practices impact student learning within particular contexts that closely resemble the actual contexts of teachers' work.

Performance Understanding

Research ers and teacher educators at Michigan State University, the University of Michigan, and elsewhere have for some time been involved in major efforts to develop professional education for prospective and experienced teachers—particularly in mathematics—that generates teaching strategies in keeping with new curriculum standards and reform-oriented pedagogies (Ball & Cohen, 1999; Lampert & Ball, 1998; Wilson & Ball, 1996; Cohen, McLaughlin, & Talbert, 1993; Cohen & Ball, 1990). Here teacher education outcomes are framed as the alignment over time of teachers' pedagogy with current curriculum standards and with discipline-based goals for students' learning of complex forms of reasoning, problem solving, and communication. This approach to performance understanding is based on earlier explorations of teachers' learning of "adventurous teaching" (Heaton & Lampert, 1993) or "teaching for understanding" (Cohen, McLaughlin, & Talbert, 1993; Cohen & Ball, 1990), conceptualized as a kind of educational practice where "students and teachers acquire knowledge collaboratively, where orthodoxies of pedagogy and 'facts' are continually challenged in classroom discourse, and where conceptual (versus rote) understanding of subject matter is the goal" (McLaughlin & Talbert, 1993). This work has received considerable attention as part of the "new professional development" (Hawley & Valli, 1999; Sykes, 1999) and/or as a "new pedagogy of teacher education" that is closely aligned with national standards for professional development and especially with visions for contemporary K-12 curricular reform (Lampert & Ball, 1998, 1999; Wilson & Ball, 1996; Ball, 1996; Ball & Cohen, 1999).

Writing specifically about performance and knowledge, Lampert & Ball (1999) argue that if teacher education is to prepare teachers for "the kind of ambitious teaching that reformers envision" (p. 39), then those who would reform teacher education will have to reconsider what it means "to know" something in teaching. They suggest that knowing means understanding in such a way that one is prepared to perform (or practice) in a given situation for which one cannot fully prepare in advance. They base this idea on David Perkins' and Howard Gardner's "performance perspective" on understanding:

> In brief, this performance perspective says that understanding a topic of study is a matter of being able to perform in a variety of thoughtful ways with the topic, for instance, to: explain, muster evidence, find examples, generalize, apply concepts, analogize, represent in a new way, and so on . . . Understanding something is a matter of being able to carry out a variety of 'performances' concerning the topic. (Perkins, 1993, p. 7, quoted in

> Lampert & Ball, 1999, p. 35)

Lampert, Ball and their colleagues advocate K-12 classrooms where children's performance understanding is the norm. Consistent with this idea, they advocate teacher education pedagogy where the performance understanding of teacher candidates is the norm. In this way K-12 curriculum and assessment, which are closely aligned with professional teaching and learning standards in the subject matter, are in turn closely aligned with teacher education pedagogy and performance assessment, which are also closely aligned with professional standards for teacher learning and professional practice. Initiatives based on these ideas attempt to provide social and organizational contexts for teacher education in which teachers work together in pairs or small groups where inexperienced teachers observe and reflect on the work of a more experienced one (Lampert and Ball, 1998).

Lampert and Ball (1998) emphasize *how* teacher candidates should know what they need to know rather than focusing on simply *what* they need to know. Based on the idea that teaching is an uncertain and indeterminate activity, they suggest that teachers learn how to construct knowledge by working in communities of practice. Teacher candidates learn by working with artifacts and records of practice, raising questions about these, connecting these to other concepts and theories, and so on. This notion of a "pedagogy of professional development" (Ball & Cohen, 1999) means presenting preservice students with various opportunities to conduct "pedagogical inquiry" (Lampert and Ball, 1998) based on artifacts and records that have been pre-catalogued and arranged in order to facilitate multiple perspectives, triangulation of interpretations, and retrieval and sorting of ideas in multiple ways.

For example, teacher candidates read or experience in a multimedia environment a more experienced teacher's records of practice and then reflect on these with the guidance of a teacher educator who may or may not be one and the same with the experienced teacher they have observed. As Lampert and Ball (1999) point out, these assessments tap into:

> …beginning teachers' capacities to analyze practice and develop hypotheses about it [and] . . . assemble portfolios of their work and to describe, justify, and analyze it. As important as what they know is their capacity to reason critically and professionally about their work. (p. 37)

The idea that the outcome of teacher education should be performance understanding—or linking what and how teachers know by working with artifacts and records of practice—is very much in keeping with assessments for beginning and experienced teachers designed by INTASC and NBPTS.

Teacher Work Samples

Western Oregon University's Teacher Work Sample Methodology (TWSM) has been in place since 1986 (Schalock & Myton, 1988) when the state of Oregon passed sweeping reforms of teacher education. These included the requirement that teacher certification programs provide evidence that teacher candidates could produce appreciable progress in the learning of all K-12 students (Cowart & Myton, 1997). With the implementation of NCATE 2000's new outcomes-based standards (NCATE, 1999), the work sample methodology—which is intended as both a vehicle for the learning of teacher candidates and a measurement system—has been receiving considerable attention (McConney, Schalock, & Schalock, 1998; Millman, 1997; Schalock, Schalock & Myton, 1998).Along these lines, the American Association of Colleges for Teacher Education (AACTE) has sponsored a series of workshops and institutes led by Western Oregon faculty to aid other teacher educators trying to develop systematic means of connecting teaching and learning (Schalock & Imig, 2000). Several other states are currently considering adopting this method.

Western Oregon's TWSM is a "complex, 'authentic' applied performance approach" to the evaluation of teacher candidates that is outcomes-based and grounded in a "context-dependent" theory of teacher effectiveness (Schalock, Schalock, & Girod, 1997, pp. 17- 18). Work samples represent teacher candidates' teaching of 3-5 week units of study developed through 8 distinct design steps from which faculty derive 7 broad categories of measure. These are used for decision making in teacher preparation and licensing as well as in research. Teacher candidates design units of instruction aligned with the desired outcomes, which are in turn aligned with Oregon's standards-based curriculum. They then assess their teaching in terms of K-12 student progress by means of the work sample method. Thus work samples provide a "rich and ready context for the evaluation of a teacher's knowledge and skill as well as a one-of-a-kind context for evaluation of teachers' effectiveness and/or productivity" (Schalock, Schalock, & Girod, 1997, p. 19).

Although the authors note that the TWSM does not stipulate specific performance standards, which are to be determined by the particular group or program using TWSM, they do provide information about how the Western Oregon program deals with evaluative criteria and performance standards. The following is illustrative of how the TWSM constructs performance as an outcome of teacher education:

> Starting with preinstructional data on pupil learning, a student teacher calculates a 'percentage correct' score for each pupil in his or her classroom. Using these scores, the teacher than (a) tabulates, from highest- to lowest-scoring

> pupil, the range of preinstructional scores; (b) sorts these scores into high-, low-, and middle-scoring groups; and (c) calculates the means scores for each of the groups formed and for the class as a whole. These preinstructional groupings provide the structure for both the analysis of postinstructional measures of outcome attainment and the calculation of gain scores.
>
> As in the case of the preinstructional measure, a percentage-correct score is calculated for each pupil on the postinstructional measure and is matched with the pupil's preinstructional score. Gain scores are then tabulated for the high-, low-, and middle-scoring groups based on the preinstructional measure. Mean gain scores also are tabulated for each of these groups and for the class as a whole to obtain a general impression of the learning gains that have been made by particular groups of pupils as a consequence of instruction received. Using these data as a point of departure, the teacher can then proceed to refine them to bring a *level of standardization* to the teacher-designed and curriculum-aligned measures of pupil learning used. This is done by calculating an Index of Pupil Growth (IPG) score for each pupil. The IPG is a simple metric devised by Millman (1981) to show the *percentage of potential growth* each pupil actually achieved. The metric is calculated as follows:
>
> $$\frac{(\text{Post \% correct}) - (\text{Pre \% correct})}{(100\% - \text{Pre \% correct})}$$
>
> Multiplying this metric by 100 results in a score than can range from –100 to +100, where a negative number represents a lower score on the posttest than on the pretest, 0 represents no change from pre- to posttest, and +100 represents a perfect score on the posttest regardless of pretest performance. A negative score is rare, with most scores falling in the +30 to +80 range. (Schalock, Schalock & Girod, 1997, pp. 22-25, emphasis in original)

Following these calculations, teacher candidates write an explanation for why K-12 students did or did not attain the desired learning outcomes. According to its architects, the teacher work sample approach to performance as outcome sharply contrasts with assessments that feature portfolios, teachers' analyses of lessons planned and taught, candidates' assessments of students' learning for diagnostic purposes, and so on. TWSM developers argue that these other approaches provide "relatively weak evidence of the teachers' success in fostering learning" (Schalock, Schalock, & Myton, 1998, p. 469) as opposed to TWSM, which focuses explicitly on demonstrable teacher effectiveness as measured by the learning gains of students.

Inquiry as Stance

For a number of years, a group of us as university- and school-based researchers and practitioners at the University of Pennsylvania and the Philadelphia area schools (and more recently at Boston College) have been involved in efforts to promote teacher research as a vehicle for generating local knowledge and challenging the status quo by linking inquiry, professional knowledge, and professional practice across the teaching lifespan (Cochran-Smith, 1991; Cochran-Smith & Lytle, 1990, 1993, 1999, 2000; Cochran-Smith , et. al., 1999). In our efforts, we have not used the language of "outcomes" and "results." However it is clear in all of the writing about these initiatives that a major outcome of teacher education is teacher learning and professional practice that promote rich learning opportunities for all students with the larger goals of equity and social justice. We have pointed this out explicitly:

> Here we take the more radical position that learning from teaching ought to be regarded as the primary task of teacher education across the professional lifespan…This argument is based in part on the assumption that the increasing diversity of America's schools and schoolchildren and the increasing complexity of the tasks that educators face render global solutions to problems and monolithic strategies for effective teaching impossible. Hence, what is required in both preservice and inservice teacher education programs are processes that prompt teachers and teacher educators to construct their own questions and then begin to develop courses of action that are valid in their local contexts and communities (Cochran-Smith & Lytle, 1993, p. 63)

From this perspective, the goals of teacher education include teacher candidates' learning to engage in practitioner inquiry and to construct local knowledge within inquiry communities (Cochran-Smith & Lytle, 1999, a; Lytle & Cochran-Smith, 1992). This work has received considerable attention as part of the teacher research movement over the last decade (Cochran-Smith & Lytle, 1990, 1999b) and has been recognized and supported nationally by the Spencer Foundation, Teachers College Press, and the University of Pennsylvania's Ethnography and Education Research Forum What professional performance looks like when inquiry is regarded as an outcome has been spelled out in detail in my writing about inquiry-centered preservice teacher education with the goal of social justice (Cochran-Smith,1991; 1995a,b; 1998) and in the writing and presentations of my students at the University of Pennsylvania and to a lesser extent at Boston College (e.g., Maimon, 1999; Black, et. al., 1993). Inquiry performances include: analyses of the culture of the school; small-scale classroom studies that drawing on classroom data, including students' written work, verbal interactions, observations, texts and other materials; case studies that explore patterns in students' classroom behavior, uses of linguistic and cultural resources, and responses to learning opportunities as well as documentation of the teacher's adaptations to these individual variations; and development of curriculum and pedagogy that provide all students (including very young children and "at risk" students) opportunities to debate complex ideas, interpret unabridged texts, exchange points of view with others based on evidence and experience, and explore issues related to equity, language, power, and racism in the classroom. These performance outcomes were developed collaboratively by university-based and school- based educators at the University of Pennsylvania over the course of many years of joint work. Fieldwork supervisors and school-based cooperating teachers had a strong voice in the development of criteria for assessment of performance, including what counted as evidence of teaching skill, students' learning, and inquiry stance. Teacher candidates were evaluated jointly—by themselves, their cooperating teachers, and their fieldwork supervisors—based on specific classroom evidence and documentation of the major goals of the program. In addition, portfolios of all teacher candidates' inquiries, samples of teachers' and students' work, and critical narrative essays analyzing teacher learning over time represented a major final performance (Cochran-Smith, 1998).

When teacher inquiry is framed as an outcome, professional performances are expected to demonstrate how teachers construct local knowledge, how they open their decision- making strategies to critique, and how they know when and what their students have learned. They also demonstrate how prospective teachers learn to wrestle with multiple perspectives, utilize others' research to generate questions and new analyses, and work within professional communities committed to social justice. Each of these aspects of learning to teach is related to what Susan Lytle and I have called an "inquiry stance" on teaching and learning (Cochran-Smith & Lytle, 1993, 1998, 1999a, 2000). Learning to teach through inquiry is difficult and uncertain work. It is work that is profoundly practical in that it is located in the dailiness of classroom decisions and actions, including teachers' interactions with their students and families, choices of materials and texts, uses of formal and informal assessments, and so on. At the same time, however it is work that is deeply intellectual in that it involves a continuous process of constructing understandings, interpretations, and questions. Performances that demonstrate that teacher candidates are learning through inquiry to teach for social justice, then, include not only the particular practices they employ and the impact these have on K-12 students' learning—but also how they struggle to document, theorize, and alter their practice.

Looking Across Constructions of Performance

The four preceding examples are similar in important ways. All four assume that a rightful outcome of teacher education is that teacher candidates can demonstrate classroom practices and accomplish classroom tasks that are linked to students' learning. All assess performance by focusing on authentic school and classroom tasks that are close to the everyday work of teaching. All assume that teacher candidates should know how to learn from their own practice by analyzing teaching and learning events and making their interpretations public and thus open to critique by others. And finally, all four make it clear that professional performance as an outcome of teacher education has to do with demonstrating the connections among teacher learning, professional practice, and student learning.

There are also important differences here, however, and the four examples provide some sense of range and variation in how professional performance is being constructed as an outcome of preservice teacher education. With approaches such as teacher work samples, for example, teacher candidates demonstrate their knowledge by constructing appropriate learning objectives and writing explanations about why particular students did and did not make the desired learning gains. In these explanations, teacher learning and teacher knowledge are regarded only as "enablers" of desired student outcomes (Schalock, Schalock, & Myton, 1998, p. 469) rather than as outcomes of teacher education themselves (Diez, 2000). The overriding focus with work samples is "demonstrable teacher effectiveness as measured by the learning gains of students" (Schalock, Schalock, & Myton, 1998, p. 469), an approach that contrasts with assessments that emphasize portfolios and inquiries by teacher candidates about students' learning, which as I stated above, are considered by work sample proponents as "weak evidence" of teacher candidates' success. In contrast to work samples, performance assessments that focus on teacher knowledge and understanding are more consistent with the professional standards of NBPTS and INTASC (Darling-Hammond, 1998; Diez, 2000). Advocates of portfolios and the like point out that teacher work samples do not provide a well-developed explanation of the connections between teaching and learning, do not require teacher candidates to understand why certain practices lead to student learning, and do not require them to justify why certain learning objectives are more important than others.

As these four examples make clear, when professional performance is regarded as an outcome of teacher education, there is variety in emphasis on teacher learning, student learning, and/or the relation between teacher and student learning. There is also variation in the sources of standards and criteria for evaluation of performances. Some of the examples above evaluate teacher candidates' performances against standards aligned with professional curriculum and teaching standards, some against standards of professional practice validated in the field, and some against some combination of these.

With other approaches, it is not clear what the sources of standards and criteria are. Along different lines, some versions of professional performance emphasize critique of curriculum standards and traditional practices by evaluating teacher candidates—at least in part—in terms of their ability to challenge, rather than comply with, current "best practice" if and when these best practices do not serve the interests of particular groups of students.

I would argue here that at the heart of different constructions of what constitutes competent teaching performance is more than a semantic debate about whether teacher education should be producing what some have called "accomplished teachers," who know how to learn from teaching on an ongoing basis, or as others have termed it, "teachers who can accomplish something" by way of measured student learning gains (Schalock & Imig, 2000). What is at the heart are basic differences in definitions of teaching and learning and in connections that are assumed among teacher learning, professional practice, and student learning. As my examples attest, these differences are played out in the tasks teacher candidates are expected to perform, the kinds of products they are required to produce, the evidence that is collected to document these, the criteria used to evaluate the evidence, and the underlying assumptions about professional knowledge and practice that guide the overall enterprise. Also at issue are the roles critique and inquiry are assumed to play (or not) in professional performance and the larger political, professional, and social agendas to which they are connected.

Constructing Outcomes in Teacher Education: Possibilities and Pitfalls

So far in this article, I have tried to make the case that how we construct outcomes in teacher education (including how we make the case that some outcomes matter more than others) legitimizes but also undermines particular points of view about the purposes of schooling, the nature of teaching and learning, and the role of the teacher in educational reform. In the remaining sections of this article, I explore some of the possibilities as well as the pitfalls in the outcomes debate.

Tensions between Consensus and Critique

Many discussions about outcomes in teacher education begin with the assumption that there is an unprecedented professional consensus about how to reform education by developing closer and closer alignment among three things: (1) standards for teaching and learning in particular content and curricular areas, (2) high stakes assessments of students and teachers, and (3) new models of teacher education, licensing, and certification. There is, however, a fair amount of evidence that just below the surface of common language and very general agreement, there are deep differences rather than consensus.

The whole movement for the privatization of schooling (and with it the deregulation of teacher education), driven by a market approach to education reform (Earley, 2000), is an obvious—an enormous—example of the lack of consensus about teacher education in the U.S. The deregulation movement mentioned earlier in this article helps to explain some otherwise puzzling discrepancies within and among state policies. For example, many states now have official relationships with NCATE and/or are working with INTASC and NBPTS to develop professional standards for the licensing of beginning teachers (Scannell & Metcalf, 2000). However some of these very same states have recently implemented or are about to put into place state policies that are fundamentally out of sync with the professional standards of these organizations. Colorado, for example, has removed the word "diversity" from its regulations regarding teacher preparation. Massachusetts Department of Education officials have excised the word "constructivism" from discussions and guidelines for school district leaders. Just two weeks before it was to be administered to thousands of K-12 students (and well after teachers and school districts had adjusted curriculum and instruction so that they would be consistent with new assessments), Arizona suspended its "cutting edge" performance-based student assessment plan and returned to more traditional assessments (Smith, Heinecke, & Noble, 1999). In addition, states such as New Jersey and Texas now advocate alternate routes with "quickie" teacher education workshops as a preferred entry into teaching (Klagholz, 2000), and new teacher certification regulations such as those in Massachusetts explicitly separate the development of pedagogy, which is to be picked up on the job, from the development of subject matter knowledge, which is regarded entirely as an arts and sciences matter (Massachusetts Department of Education, 1999).

These are glaring examples of the fact that there is not consensus in the U.S. about how and where teachers should be educated, what they should learn (or not learn), and what theories of teaching and learning should guide their learning. Even if we put the professionalization- deregulation debate aside, however, it may be that what Hawley and Valli (1999) have called "an almost unprecedented consensus . . . among researchers, professional development specialists, and key policymakers on ways to increase the knowledge and skills of educators substantially" is at least partly an illusion—or a wish.

There are indications of lack of consensus within the profession as well as between the profession and its detractors. For example, only 500 of the 1200 institutions in the country that recommend teachers for certification are nationally accredited (Wise, 1999), and Linda Darling-Hammond (2000) claimed in a recent discussion of the reforms called for by the NCTAF that the American Association of Colleges for Teacher Education had actually lobbied against a provision in the Higher Education Act that would have encouraged accreditation as a means of increasing accountability for teacher education institutions. (Note 3) Along related but different lines, Frank Murray, who was an early and active player in efforts to codify the knowledge base for teaching and teacher education (Murray, 1996), has cautioned that the knowledge base is a tentative and emerging one with few settled policies and practices (Murray, 2000). He points out that the professional standards, which are the backbone of reforms proposed by NCTAF and other professional agencies, represent provisional and untested recommendations rather than empirically validated policies and practices. Murray advocates accreditation standards based on outcomes evidence in keeping with institutional purposes and goals rather than simply in keeping with standards. Murray and the TEAC organization, which he heads, have been characterized as obstacles to reform in teacher education, and their emphasis on outcomes evidence based on institutional goals rather than professional standards has been labeled "disingenuous" at best, "consumer fraud" at worst (Darling-Hammond, 2000, a).

Along different lines, Susan Lytle and I have argued (Cochran-Smith & Lytle, 2000) that the widely touted "new professional development" may be less monolithic and consensual than is claimed in some places. We have suggested that beneath the surface of similarly- named teacher education strategies and organizational arrangements such as professional development schools or inquiry-centered teacher education, "the new vision" of professional development differs substantially, depending in part upon underlying assumptions and goals, especially upon differing images of knowledge, practice, and teacher learning (Cochran-Smith & Lytle, 1999, a).

Some of the differences noted above among teacher education policy makers, researchers, and practitioners may be accounted for as turf battles, some as what Smith, Heinecke, and Noble (1999) call "political symbolism and contention" (p. 158), and some as genuine and rational debate about the meaning of teaching and learning and the purposes of schooling. But in the face of these disagreements, it is appropriate to ask what accounts for the strong claims that consensus already exists and what propels such strong advocacy of closer and closer alignment of educational outcomes.

Yinger's incisive explanation of the role of standards and consensus in the process of professionalization (Note 4) is useful here (Yinger, 1999; Yinger & Hendricks-Lee, 2000). He points out that the central issue in professionalization is how a group makes a claim for and establishes "jurisdictional authority" (Yinger, 1999, p. 86) over the knowledge and problems of professional practice in a given area. He comments that standards are a powerful professional tool and that consensus is critical to the professionalization process, signaling to the public and to policy makers that a profession has established cognitive jurisdiction. Yinger concludes:

> As consensus develops around national standards for teaching and teacher preparation, it fulfills the needs of both policy makers and the public for simplification of the image of teaching and issues of quality. There was no way teaching could have met these social needs for a unified, scientifically based perception of professional practice as

> long as academics were arguing publicly about conceptions of teaching and 50 state legislatures were deciding the matters for themselves. (p. 106)

Yinger's analysis suggests that we need consensus about outcomes in teacher education whether we have it or not. The pitfall here—and my caution as we construct outcomes in teacher education—is that we will sacrifice or gloss over the healthy and vital contribution of critique for what is arguably the greater professional good of consensus.

On a certain level, working from consensus and alignment of standards at multiple levels of schooling and teaching are rational and much-needed improvements in teacher education. Aligning school-based curriculum and learning standards with standards for teacher education is a far cry from the days of haphazard or idiosyncratic teacher education programs based on faculty members' favorite assignments or distant memories of their own teaching experiences. On another level, however, the greater the supposed consensus and the tighter the alignment of all the pieces, the less room there is for critique and questioning within the profession and in the preparation of prospective teachers.

As we construct outcomes in teacher education, a central challenge is how to prepare teacher candidates who can demonstrate what some consider "best" instructional practices, but also know how to challenge those practices when they exclude certain children or fail to serve some students. How will we prepare teachers who know how to "fit" into tightly aligned standards- driven schools and school systems, but also know how to raise questions about whose interests are being served, whose needs are being met, and whose are not being met by those systems?

The emerging professional consensus is that teacher candidates must demonstrate that they can affect the learning of *all* K-12 students. But serving the needs of *some* K-12 students may mean challenging the consensus itself—challenging the bases of some curriculum frameworks, assessments, and school policies that do not serve all students by identifying inequities in the current arrangements of schooling. Critique as an outcome of teacher education—"teaching against the grain" as outcome (Cochran-Smith, 1991a)—is a notion that is diametrically opposed to recent initiatives in some higher education institutions that are intended to provide "quality assurances" about their recent graduates. Quality assurances, or warranties—if you will—are commitments made by higher education institutions to local school districts that if their teacher candidates, once hired, are not able to perform to the satisfaction of school principals on their first jobs, they will be assisted and "retrained" by the teacher education institution until they can. What does this kind of quality assurance do to the notion of the "learning teacher" who teaches to standards but also critiques them? What does this do the notion of teacher as professional decision-maker who faces difficult choices among competing claims to justice in order to meet the needs of all students? In teacher education, we face a major challenge—how to retain and nurture constructive critique at the same time that we work to build professional consensus about what makes a promising teacher candidate and a good teacher.

Problems with the Inputs-Outputs Metaphor

As mentioned above, some people have been describing changes in accreditation standards as a "paradigm shift" (Schalock & Myton, 1988; Schalock & Imig, 2000) from "inputs to outputs" or from "inputs to outcomes" in teacher education. It is certainly appropriate to acknowledge that there are major differences in NCATE's new accreditation standards and in the new general focus on results and outcomes. NCATE's new standards focus less on the knowledge bases and conceptual frameworks of teacher education programs and more on systematic evaluation of teacher candidates' demonstrated ability to foster K-12 students' learning (NCATE, 1999). It is also the case that from its inception, TEAC focused on outcomes rather than inputs—that is, TEAC's approach was from the beginning a system for auditing the performances of teacher candidates and programs rather than assessing the alignment of curricula and programs with professional standards (TEAC, 1999).

There are a number of problems, however, with characterizing this change in emphasis as a paradigm shift and in using metaphors such as "inputs and outputs" to describe it. In Kuhn's sense, the phrase, paradigm shift, implied a major C change and a major change in world view that was shared by a given research or academic community. To apply the paradigm shift phrase to new and old ways of accrediting teacher education programs implies at the very least, that "old" programs—those that focused on the "inputs" of teacher education courses and curriculum—had nothing to do with teacher candidates' actual teaching or with K-12 students' actual learning and that old programs had little concern with how teacher candidates adjusted their professional practice to meet the needs of diverse learners. As many teacher education practitioners and researchers are well aware, however, this is not the case.

There have been many programs over the last two decades that have had all along what we might now call an "outcomes" focus, particularly those that were inquiry- and/or research-based, those that were situated within the ongoing work of schools and classrooms, and those that were committed to preparing teachers for urban and special needs populations. These programs have long concentrated on how teacher candidates posed questions, documented students' learning, analyzed and interpreted classroom data, adjusted the curriculum to meet the needs of different students, and critiqued their own and others' practice. (Note 5) Characterizing new accreditation standards as a "paradigm shift" fails to acknowledge that programs like these have long emphasized learning to teach as a process of learning to document systematically teachers' and students' learning.

However, the dominance of the input-output metaphor to describe teacher education outcomes is even more troubling than overuse of the paradigm shift phrase. The input-output metaphor conjures up production and factory imagery and calls to mind the linear flow charts of early computer programming days and the schematics that were used to represent the input- output operations of early technology. In *Metaphors We Live By*, Lakoff and Johnson (1980) suggest that images like these can be powerful forces in the social construction of reality:

> Metaphors may create realities for us, especially social realities. A metaphor may thus be a guide for future action. Such actions will, of course, fit the metaphor. This will, in turn, reinforce the power of the metaphor to make experience coherent. In this sense metaphors can be self-fulfilling prophecies. (p. 156)

The input-output metaphor carries with it a linear view of the relationship of teaching and learning for both K-12 students and for teacher candidates, an image that is somewhat reminiscent of the process-product research that dominated research on teaching not so long ago (Dunkin & Biddle, 1974). With process-product research, teacher behaviors were central. Teacher education programs consistent with this research base made certain their teacher candidates could demonstrate these behaviors in classroom settings. In current constructions of the outcomes question, there is a different focus—a focus on K-12 student learning rather than teacher behaviors. Schalock, Schalock, and Girod (1997) points out explicitly that the new focus on outputs and results is quite different from process-product approaches in that the contexts of teaching are acknowledged and the emphasis is on student learning as opposed to teacher behaviors. Despite these differences between process-product research and outcomes-based evaluation of teacher education, however, their underlying conceptions of teaching and learning are similar—and linear—as the input-output metaphor so powerfully suggests.

As we construct outcomes for teacher education, an important challenge will be to eschew narrow views of teaching, particularly those that begin and end with the assumption that teaching can be defined as instructional practice that leads to demonstrable student learning gains. If we require teacher candidates to use some kind of calculus that measures and aggregates the learning gains of each K-12 student from pretest to posttest measures for each lesson or teaching unit, there will be an inevitable narrowing of the curriculum and an inevitable pull toward teaching as transmission and learning as accruing bits of knowledge. There will also be an inevitable emphasis on teaching practice as what teachers do within the boundaries of their classroom walls rather than an expanded view that includes teachers' roles as members of school communities, activists, school leaders, and theorizers of practice. I have described this broader view of teaching practice as follows (Cochran-Smith & Lytle, 1999, a):

> This image of practice entails expanded responsibilities to children and their families, transformed relationships with teachers and other professionals in the school setting, as well as deeper and altered connections to communities, community organizations, and school-university partnerships. We are not suggesting that an expanded view of practice results from adding teachers' activity outside the classroom to what they do inside, but rather that what goes on inside the classroom is profoundly altered and ultimately transformed when teachers' frameworks for practice foreground the intellectual, social, and cultural contexts of teaching (p. 276).

In short, what I am suggesting here is that we need outcomes measures that—ironically—make teaching harder and more complicated for teacher candidates (rather than easier and more straight-forward). Such measures recognize the inevitable complexity and uncertainty of teaching and learning and acknowledge the fact that there are often concurrent and competing claims to justice operating in the decisions teacher candidates must make from moment to moment, day to day. Linear models of teaching will not suffice here, nor will constructions of outcomes that push only for clarity and certainty. Someone once said that those who have been forced to memorize the world are not likely to change it. It may also be true that those who have been required to measure the outcomes of teaching only with pluses and minuses will not be likely to see the value of question marks, concentric circles, and arrows that point both ways and sometimes double back.

Teachers (and Teacher Educators) as Saviors and Culprits

Many of the outcomes discussions in teacher education are based on the premise that teachers and teaching, teacher educators and teacher education, are critical components—arguably *the* critical components—in school change (and ultimately perhaps societal change). There is good news and bad news here. In debates about outcomes, teachers and teacher educators are being constructed as both the last great hope and the most culpable culprits in what ails American schools, a point that has been made repeatedly, often using quotations like these from Michael Fullan and David Cohen, respectively:

> Teacher education still has the honor of being simultaneously the worst problem and the best solution in education. (Fullan, 1993, p. 105 quoted in Thiessen, 2000, p. 129)

> Teachers are the problem that policy must solve, in the sense that their modest knowledge and skills are one important reason why most instruction has been relatively didactic and unambitious. But teachers are also the agents on whom policy must rely to solve that problem, for unless they learn much more about the subjects they teach, and devise new approaches to instruction, most students' learning will not change. (Cohen, 1995, p. 13 quoted in Schalock & Imig, 2000, p. 6)

The attention given recently to outcome-based assessment systems that incorporate student achievement data into evaluations of individual teachers and schools reinforces this idea. The research of Sanders and Horn (1994, 1998), for example, based on their Tennessee Value-Added Assessment System has been widely cited by researchers and policy makers who represent a wide range of perspectives (e.g. Darling-Hammond,1998, 2000; Murray, 2000; Ballou & Podgursky, 1999) and even reach diametrically different conclusions about teacher education and teacher licensing policies. Despite their differences, however, policy makers use research like Sanders and Horn's to make the same point about the importance of teachers and teachers' work: When other variables are adjusted for or held constant, teacher effectiveness is the primary factor that accounts for differences in student learning, even stronger as a determinant of students' achievement than class size and heterogeneity. This means that teachers are responsible for students' learning despite the mitigation of social and cultural contexts, students' backgrounds, and the match or mismatch of school and community expectations.

Many of the most prominent voices in discussions about outcomes use evidence about the impact of individual teachers to make an equally strong point about the importance of teacher education. This link is crystal clear in Gary Sykes' (1999) introduction to a recent handbook of policy and practice, which he co-edited with Linda Darling-Hammond (Darling-Hammond & Sykes, 1999).

> Improvement of American education relies centrally on the development of a highly qualified teacher workforce imbued with the knowledge, skills and dispositions to encourage exceptional learning in all of the nation's students. (Sykes, 1999, p. xv)

My intention here is *not* to differ with Sykes and others who are adamant about the importance of teacher professionalization. I am in no way suggesting that teachers—and teacher education—are not important. I have spent more than twenty years demonstrating and acting on the assumption that they are. During this time, I have argued consistently that we need teachers who enter and remain in the profession not expecting to carry on business as usual but prepared to teach differently and to join others in major efforts to change the ways we think about teaching, schooling, and social change (Cochran-Smith, 1991, 1995b, 1998).

As we construct outcomes for teacher education, we face the challenge of how to emphasize the centrality of teachers' work without implying that teachers—individually or collectively—are the panacea for the problems of American education and American society. The dire circumstances of the cities are not going to change because teachers teach better. Weiner (1989) makes this point with clarity when she argues that the "Herculean task" of teaching in urban schools is the result of complex school bureaucracies, the isolation of schools from the families and communities they are supposed to serve, and the large numbers of students in urban classrooms whose families have neither the resources nor the will to affirm and support school values. Weiner points out that professional development projects can only help teachers deal with the third factor—the situations they find in their classrooms:

> Teacher education programs can prepare teachers to confront ...conditions in their classrooms, by educating candidates to teach disadvantaged students with respect, creativity, and skill, but they cannot prepare individual teachers to substitute for the political and social *movements* that are needed to alter the systemic deficiencies of urban education. (p. 153)

McCarthy (1993) makes a similar point in his criticism of multicultural education. He claims that by ignoring "the crucial issues of structural inequality and differential power relations" (p. 243), advocates of multicultural education place enormous and unrealistic responsibility on the shoulders of classroom teachers. Notwithstanding recent research about the enduring impact of teacher expertise on students' learning, we must remember that teachers—and teacher educators—are neither the saviors nor the culprits of all that is wrong with American education and American society.

Getting Social Justice onto the Outcomes Agenda

In the standards of NBPTS, INTASC, and NCATE, there is an explicit mandate that teachers and teacher candidates meet the needs of an increasingly diverse student population by producing demonstrable learning gains for *all* children. NBPTS Standard 1 states that professional teachers must be committed to students' learning and dedicated to making knowledge accessible to all students and that expert teachers adjust their teaching according to varying student interest, skill, knowledge and background (National Board for Professional Teaching Standards, 1994). Similarly INTASC Principle 3 states that the good beginning teacher understands "how students differ in their approaches to learning and creates instructional opportunities that are adapted to diverse learners" (Interstate New Teacher Assessment and Support Consortium, 1992). NCATE's new Standard 4, which is labeled "Diversity," is consistent with NBPTS and INTASC standards. It requires that teacher preparation units must design, implement, and evaluate curriculum, field experiences, and clinical practices so that teacher candidates acquire the knowledge, skills and dispositions necessary to help all students learn. NCATE stipulates that this should include experiences working with diverse higher education and school faculty, diverse teacher candidates, and diverse and exceptional students in schools (National Council for the Accreditation of Teacher Education , 1999). In particular, NCATE standards require that "candidates learn to contextualize teaching and to draw upon representations from the students' own experiences and skills. Candidates should learn how to challenge students toward cognitive complexity and engage students through instructional conversation" (pp. 15-16).

Some proponents of teacher professionalization have pointed out that the standards of NBPTS and INTASC coupled with new NCATE standards provide a remarkably consistent picture of the good teacher. Yinger (1999) makes this point quite lucidly:

> Through the work of [these] three organizations…a powerful consensus has emerged regarding the definition and assessment of good teaching throughout a career, from preservice education to advanced professional certification. The standards have framed the image of the professional teacher as a knowledgeable, reflective practitioner willing and able to engage in collaborative, contextually grounded learning activities. (p. 102-103)

An image of the professional teacher as reflective and knowledgeable is certainly laudable, one that few would debate. It is also important to ask, however, whether this emerging view of the prospective professional includes images of teacher candidates as activists, as agents for social change, and/or as allies for social justice? Does it include an image of the teacher candidate as one who works with others to challenge the current arrangements of schools and schooling?

As we construct outcomes in teacher education, we need to interrogate what it means to teach "all students" well and what it means to adjust teaching practices according to the needs and interests of "all children." In a recent chapter on preparing teachers for diversity, Gloria Ladson-Billings (1999a) asserts that "the changing demographics of the nation's schoolchildren have caught schools, colleges, and departments of teacher education by surprise. Students are still being prepared to teach in idealized schools that serve White, monolingual, middle class children from homes with two parents" (p. 86-87). In another recent article about culturally relevant approaches to teacher assessment, Ladson-Billings (1999b) further asserts that these are "dangerous times" for teachers of students of color because some of the new evaluations of teacher competency "may actually serve to reinscribe a narrow set of teaching practices that fail to serve all children well—particularly children of color and children living in poverty" (p. 255). Similarly Jackie Jordan Irvine suggests that some aspects of current teacher assessments, including those used by NBPTS, are not in keeping with what we know about the strategies, relationships, and beliefs of teachers who teach children of color most effectively (Irvine, 2000; Irvine & Fraser, 1998).

As we construct outcomes in teacher education, one of the challenges we face is how to keep social justice—particularly issues of race, class, and language background—on the agenda. At the same time that there is a professional consensus that the professional teacher is knowledgeable, reflective, and collaborative, another consensus has emerged about the effective teacher of children of color, children whose first language is not English, and/or children whose culture is not Western European in origin. This other image of the professional teacher is of one who constructs pedagogy that is culturally relevant and responsive (Gay, 2000; Irvine & York, 1995; Ladson-Billings, 1994, 1995), multicultural but also socially reconstructionist (Sleeter & Grant, 1987; Sleeter & McLaren, 1995), anti-racist (Sleeter, 1992; Tatum, 1992), anti-assimilationist (King, 1996), and/or aimed at social justice (Cochran-Smith, 1995, a,b; 1999). (Note 6) In short, the professional teacher is one who teaches in a way that bell hooks (1994) calls emancipatory or "transgressive":

> The classroom with all its limitations, remains a location of possibility. In that field of possibility we have the opportunity to labor for freedom, to demand of ourselves and our comrades an openness of mind and heart that allows us to face reality even as we collectively imagine ways to move beyond boundaries, to transgress. This is education as the practice of freedom. (p. 207)

I want to be clear that I am in no way suggesting that these two images of the professional teacher—as reflective and knowledgeable, on the one hand, and as transformative and culturally relevant, on the other—are necessarily inconsistent or that they cannot mutually coexist in constructions of outcomes in teacher education. In fact with performance assessments where teacher candidates are expected to document student learning but also demonstrate their own efforts to work for social change, the two images are entirely consistent and mutually reinforcing. But it is also important to note that these two images are by no means *necessarily* co-incidental. We could easily imagine performance assessments, for example, that demonstrate that a teacher candidate is reflective, collaborative, and knowledgeable but that have little or nothing to do with critiquing the inequities of the educational system or raising questions about the school as a sorting machine that reinforces privilege as well as disadvantage. An important challenge as we construct outcomes for teacher education is to imagine performance assessments for teacher candidates that require both.

Outcomes in Teacher Education: Democratic or Market Driven?

As I have alluded several times, many of the most contentious debates about outcomes in teacher education stem from two fundamentally different approaches to teacher education reform and from two fundamentally different views of the purposes of schooling. The first, which is intended to reform teacher education through professionalization so that all students are guaranteed fully-licensed and well- qualified teachers, is based on the belief that public education is vital to a democratic society. The second, which is intended to reform teacher education through deregulation so that larger numbers of college graduates (with no teacher preparation) can enter the profession, is based on a market approach to the problem of teacher shortages that feeds off erosion of public confidence in education.

A number of analysts have argued that a market approach to educational policy fundamentally undermines a democratic vision of society (Earley, 2000; Engel, 2000; Labaree, 1997). Michael Engel (2000) makes this point bluntly: "Market ideology and democratic values in education are mutually exclusive" (p. 6). Similarly Earley (2000) and Labaree (1997) each point out that a market approach to reform of teaching and teacher education fundamentally misunderstands the nature of teachers' work, which is primarily a public enterprise for the common good, in contrast with market approaches to educational reform, which are about individual competition for what Labaree calls "private goods." Pointing to some of the basic contradictions implicit in the 1998 Higher Education Act as evidence of the mismatch between teachers' work, which is fundamentally democratic, and market-driven reforms, which are fundamentally competitive and individualistic, Earley offers this trenchant analysis:

> A market policy lens is based on competition, choice, winners and losers, and finding culprits. Yet teachers must assume that all children can learn, so there cannot be winners and losers. Market policies applied to public education are at odds with collaboration and cooperative approaches to teaching and learning…Paradoxically the Higher Education Act Title II categorical programs encourage institutions of higher education to form collaborative partnerships across academic disciplines and with K-12 schools for the purpose of preparing new teachers and offering professional development for career educators. However, under the market approach being used in educational policy and reflected in the

> accountability sections of the same law, teachers and those who design and administer their preparation programs must have as a primary concern competition, being a winner, not a loser, and certainly not being cast as a culprit. The consequence of these pressures is the domestication of teachers (Note 7) [and perhaps I could add here, the domestication of teacher educators], perpetuating their role as semiskilled workers. . . and frustrating efforts for teaching to be truly professional work. (pp. 36-37) [parenthetical comment added]

Constructions of outcomes that are embedded within market approaches to education reform legitimize the dominance of "private goods" and undermine the view that public education is an enterprise for the public good in a democratic society. Emphasis on private goods and the privatization of education is a trend that is not limited to the U.S. Rather the free-market approach to educational reform is a global phenomenon. Along these lines, Apple (2000), Whitty, Power, & Halpin (1998), and Robertson (1998), among others, have pointed out that the tendency in Australia, New Zealand, the U.K., and in parts of the U.S. has been to devolve blame for the "failures" of public education to the local level—schools, teachers, and teacher education programs—while at the same time over- regulating the content of education and dramatically curtailing the role of universities in teacher education (Thiessen, 2000).

Many of the recent attacks on teacher education are best understood in terms of this lager global debate. There is a striking similarity in many of the attacks on teacher education and in their allegiance to market-driven reforms that make the anti-democracy theme very clear . In these attacks, multicultural education is often constructed as a villain (Farkas & Johnson, 1997; Schrag, 1999)—at best politically correct but meaningless, and at worst an evil political movement that is denying white middle class citizens their share of space in the pages of textbooks and causing a downward trend in children's skills (Stotsky, 1999). In many of the attacks on teacher education, the commentator presumes to speak for "the public," for "public school teachers," or for "parents," all of whom want the same things—order, discipline, basic skills, and a return to American traditions (Farkas & Johnson, 1997). There is also an assumption that knowledge is a static and inert commodity that is (or should be) transmitted directly from teachers to students. Finally there is the presumption that what would save our schools is the "return" to an earlier and idealized time when American values were uncontested and shared by all, when the "canon" of western European history and literary works was unchallenged, and when academic standards for all students were rigorous and culturally neutral (Ravitch, 2000). Each of these entirely faulty presumptions and historical inaccuracies has been critiqued and deconstructed in great detail elsewhere (e.g., Apple, 2000; Banks, 2001; Ladson-Billings, 1999a).

The similarities among many of these attacks, though, are not surprising—nor are their explicitly conservative politics and their gestures toward racism—when it is understood that they are part of a market-driven approach to educational reform and part of the larger conservative political agenda for the privatization of American education. Although it claims to be neutral, this agenda begins with the premise that we need to deregulate and dismantle teacher education, certifying teachers solely on the basis of high stakes test scores and letting the market decide which children will have the most qualified teachers. These are anything but neutral premises and neutral assumptions about the purposes of American education, the purposes of teacher education, and the role of public education in a democratic society.

Mary Heaton Vorse once wrote, "In the last analysis, civilization itself will be measured by the way in which children live and by what chance they have in the world" (quoted in Maggio, 1997, p. 8). As we construct outcomes for teacher education, we need to keep in mind how *we* will be measured by our own measures. As researchers, practitioners, and policy makers in teaching and teacher education, *we* will not measure up unless we preserve a place for critique in the face of consensus, unless we keep at the center of teacher education rich and complex understandings of teaching and learning that are not easily reducible to algorithms, unless we acknowledge that although teachers have a critical role in educational reform, they alone are neither the saviors nor the culprits in what is wrong with American schools and American society, and unless we remain vigilant in demanding time and space on the outcomes agenda not just for professional discussions about meeting the needs of all students but for deep interrogation of questions related to diversity, equity, access, and racism.. At this critical juncture in the reform and development of teacher education, if we do not take control of framing the outcomes in teacher education, then the outcomes will surely frame us and undermine our work as teachers, teacher educators, researchers, and policy makers committed to a democratic vision of society and to the vital role that teachers and teacher educators play in that vision.

Notes

The author wishes to acknowledge the insightful comments on early drafts of this paper from: Susan Lytle, Larry Ludlow , Curt Dudley-Marling, and Mary Kim Fries, who also provided invaluable bibliographic and research assistance

A version of this paper was presented as the AERA Vice Presidential Address for Division K (Teaching and Teacher Education) at the AERA Annual Meeting in New Orleans, April, 2000.

1. The American Education Research Association's "National Consensus Panel on Teacher Education" is currently exploring the empirical research in several areas related to teacher

qualifications, program structures, teacher attrition, and career choices. Part of the task of this panel is to consider contradictory claims in these areas.

2. The examples used here are drawn exclusively from preservice teacher education; thus I have not used as examples the performance assessments developed as part of early licensing requirements in various states (e.g., INTASC efforts in Connecticut, Indiana, etc.). It is important to note also that I am *not* proposing a typology of performance assessments in preservice education nor am I offering these examples as prototypes. I am also not suggesting that these are mutually exclusive from one another since they are clearly not and in fact several of them overlap or are consistent in important ways. Rather I believe that they provide some sense of the ways the performance is being constructed as an outcome in preservice education as well as some sense of the consequences of doing so.
3. David Imig, President of the American Association of Colleges of Teacher Education, suggests this characterization of AACTE's position is misleading if not inaccurate because it does not fully take into account the political issues that swirled around these debates nor the fact that there was no realistic possibility that this provision would have become policy (Imig, personal communication, 2000).
4. Yinger (1999) draws on Andrew Abbott's sociological analysis of professionalization across European and American modern professions for his analysis of professionalism and standards in teacher education.
5. See Cochran-Smith & Lytle (1999b) for a synthesis of the teacher research movement over the last ten years and Cochran-Smith Lytle (1999a) for an overview of teacher education initiatives wherein new and experienced teachers work together to construct local knowledge of practice.
6. I have argued elsewhere (Cochran-Smith, 1999) that although these various pedagogies are not synonymous, they are animated by several shared premises that comprise the idea of teaching for social justice. Schools (and how "knowledge," "curriculum," "assessment," and "access" are constructed and understood in schools) are not neutral grounds but contested sites where power struggles are played out. The structural inequities embedded in the social, organizational, and financial arrangements of schools and schooling help to perpetuate dominance for dominant groups and oppression for oppressed groups. Power, privilege, and economic advantage and/or disadvantage play major roles in the school and home lives of students whether they are part of language, cultural, or gender majority groups or minority groups in our society . The history of racism and sexism in America and the ways "race" and "gender" have been constructed in schools and society are central, whether consciously or not, in the ways students, families, and communities make meaning of school

phenomena as well as how they interact with school designates. Curriculum and instruction are neither neutral nor natural. The academic organization of information and inquiry reflects contested views about what knowledge is of most value; part of the curriculum is what is present or absent as well as whose perspectives are central or marginalized, and whose interests are served or undermined. The social and organizational structures of instruction, including classroom and other discourse patterns, grouping strategies, behavioral expectations, and interpretive perspectives are most congruent with White mainstream patterns of language use and socialization and are more conducive to the achievement of boys than girls. Animated by these understandings, teaching for social justice is teaching that is openly committed to a more just social order (Freire, 1970; Nieto, 1996).

7. Earley attributes this phrase to Philadelphia School District teacher and researcher, Diane Waff.

References

______________. *The 2000 campaign: Transcript of debate between Vice President Gore and Governor Bush*(2000, October 4)., [online website - New York Times Archives].

Alverno College. (1996). *Ability-based learning program: Teacher education*. Milwaukee, WI: Alverno College Institute.

Apple, M. (2000). Can critical pedagogies interrupt rightists policies? *Educational Theory, 50*(2), 229-254.

Ashton, P., & Crocker, T. (1987). Systematic study of planned variations: The essential focus of teacher education reform. *Journal of Teacher Education, 38*, 2-8.

Ball, D. (1996, March). Teacher learning and the mathematics reforms: What we think we know and what we need to learn. *Phi Delta Kappan, 77*(7), 500-508.

Ball, D., & Cohen, D. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3-32). San Francisco, CA: Jossey Bass.

Ballou, D., & Podgursky, M. (1997). Reforming teacher training and recruitment. *Government Union Review, 14*(4), 1-53.

Ballou, D., & Podgursky, M. (1997). *Reforming teacher training and recruitment: A critical appraisal of the recommendations of the*

National Commission on Teaching and America's Future, [online website]. http://www.psrf.org/doc/v174_art.html.

Ballou, D., & Podgursky, M. (1999). *Reforming teacher preparation and licensing: What is the evidence?* (ID: 10418), [online website]. http://www.tcrecord.org.

Ballou, D., & Podgursky, M. (1999). Teacher training and licensure: A layman's guide. In M. Kanstoroom & C. Finn (Eds.), *Better teachers, better schools*. Washington, DC: The Thomas Fordham Foundation.

Ballou, D., & Podgursky, M. (2000). Reforming teacher preparation and licensing: What is the evidence? *Teachers College Record, 102*(1), 5-27.

Banks, J. (1996). *Multicultural education, transformative knowledge and action: Historical and contemporary perspectives*. New York, NY: Teachers College Press.

Banks, J. (2001). Citizenship education and diversity: Implications for teacher education. *Journal of Teacher Education, 52*(1), 5-16.

Barnes, H. (1989). Structuring knowledge for beginning teaching. In M. Reynolds (Ed.), *Knowledge base for the beginning teacher* (pp. 13-22). New York, NY: Pergamon Press.

Black, L., Bousel, H., Byer, L., Cimakasky, L., Coy, D., Freilich, P., Hartman, B., Hilton, D., Joe, S., Lawrence, D., Mahoney-Hanley, M., Swenson, J., & Winkelstein, B. (1993). Leaving the script behind. In M. Cochran-Smith & S. Lytle, *Inside/Outside: Teacher research and knowledge* (pp. 194-203). New York, NY: Teachers College Press.

Blackwell, P., & Diez, M. (1999). *Achieving a new vision of master's education for teachers*. Washington, DC: National Council for the Accreditation of Teacher Education.

Borrowman, M. (1956). *The liberal and technical in teacher education: A historical survey of American thought*. New York, NY: Teachers College Press.

Carnegie Forum on Education and the Economy. (1986). *A nation prepared: Teachers for the 21st century*. New York, NY: Carnegie Corporation.

Clark, C., & Peterson, P. (1986). Teachers' thought processes. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 255-296). New York, NY: Macmillan.

Cochran-Smith, M. (1991). Learning to teach against the grain. *Harvard Educational Review, 51*(3), 279- 310.

Cochran-Smith, M. (1995a). Color blindness and basket making are not the answers: Confronting the dilemmas of race, culture, and language diversity in teacher education. *American Educational Research Journal, 32*(3), 493- 522.

Cochran-Smith, M. (1995b). Uncertain allies: Understanding the boundaries of race and teaching. *Harvard Educational Review, 65* (541-570).

Cochran-Smith, M. (1998). *Inquiry as a stance on teaching: The preservice case.* Paper presented at the Annual Meeting for the American Educational Research Association, San Diego, CA.

Cochran-Smith, M. (2000a). Editorial: The questions that drive reform. *Journal of Teacher Education, 51*(5), 331-333.

Cochran-Smith, M. (2000b). The future of teacher education: Framing the questions that matter. *Teacher Education Journal, 11*(1), 13-24.

Cochran-Smith, M. (in press). The outcomes question in teacher education. *Teaching & Teacher Education.*

Cochran-Smith, M., & Dudley-Marling, C. (In press). The flunk heard 'round the world. *Teaching Education.*

Cochran-Smith, M., & Lytle, S. (1990). Research on teaching and teacher research: The issues that divide. *Educational Researcher, 19* (2), 2-11.

Cochran-Smith, M., & Lytle, S. (1993). *Inside/Outside: Teacher research and knowledge.* New York, NY: Teachers College Press.

Cochran-Smith, M., & Lytle, S. (1999a). Teacher learning in professional communities. In D. P. Pearson & I. Iran-Nejad (Eds.), *Review of Research in Education* (Vol. 24). Washington, DC: American Educational Research Association.

Cochran-Smith, M., & Lytle, S. (1999b). Relationship of knowledge and practice: Teacher learning in communities. In A. Iran-Nejad & C. D. Pearson (Eds.), *Review of research in education* (Vol. 24, pp. 249-306). Washington, DC: American Educational Research Association.

Cochran-Smith, M., & Lytle, S. (1999). The teacher research movement: A decade later. *Educational Researcher, 28*(7), 15-25.

Cochran-Smith, M., Dimattia, P., Dudley-Marling, C., Freedman, S., Friedman, A., Jackson, J., Jackson, R., Loftus, F., Mooney, J., Neisler, O., Peck, A., Pelletier, C., Pine, G., Scanlon, D., & Zollers, N. (1999, March). *Seeking social justice: A teacher education faculty's self study, Year III.* Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada.

Cochran-Smith, M., Diez, M., Imig, D., Murrell, P., Pankrantz, R., & Schalock, D. (2000). *Shifting the paradigm: Professionalizing teaching through accountability for student progress in learning.* Symposium held at the American Association of Colleges for Teacher Education, Chicago, IL.

Cohen, D., & Ball, D. (1990). Policy and practice: An overview. *Educational Evaluation and Policy Analysis, 12*(3), 347-353.

Cohen, D., McLaughlin, M., & Talbert, J. (Eds.). (1993). *Teaching for understanding: Challenges for policy and practice.* San Francisco, CA: Jossey-Bass Publishers.

Cowart, B., & Myton, D. (1997). The Oregon teacher work sample methodology: Rationale and background. In J. Millman (Ed.), *Grading teachers, grading schools* (pp. 11-14). Thousand Oaks, CA: Corwin Press.

Darling-Hammond, L. (1991). Are our teachers ready to teach? *Newsletter of the National Council for Accreditation of Teacher Education, 1*, 6-10.

Darling-Hammond, L. (1996). Who teaches and why: Dilemmas of building a profession for twenty-first century schools. In J. Sikula & T. J. Buttery & E. Guyton (Eds.), *Handbook of research on teacher education* (2nd ed., pp. 67-93). New York, NY: Macmillan.

Darling-Hammond, L. (1998). Teacher learning that supports student learning. *Educational Leadership, 55*(5), 6- 11.

Darling-Hammond, L. (1999, December). State teaching policies and student achievement. *Teaching Quality: Policy Briefs*(2), 1-7.

Darling-Hammond, L. (2000a). Teaching for America's future: National commissions and vested interests in an almost profession. *Educational Policy, 14*(1), 162-183.

Darling-Hammond, L. (2000b). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives, 8*(1).

Darling-Hammond, L., & McLaughlin, M. (1999). Investing in teaching as a learning profession: Policy, problems, and prospects. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 376-412). San Francisco, CA: Jossey Bass.

Darling-Hammond, L., & Sykes, G. (Eds.). (1999). *Teaching as the learning profession: Handbook of policy and practice*. San Francisco, CA: Jossey Bass.

Darling-Hammond, L., Wise, A., & Klein, S. (1999). *A license to teach: Raising standards for teaching*. San Francisco, CA: Jossey Bass.

Denton, J., & Lacina, L. (1984). Quantity of professional education coursework linked with process measures of education. *Teacher Education and Practice, 1*, 39- 46.

Diez, M. (1996). Who will prepare the next generation of teachers? In L. Kaplan & R. Edelfelt (Eds.), *Teachers for a new millennium: Aligning development, national goals, and high standards for all students* (pp. 20-35). Thousand Oaks, CA: Corwin Press.

Diez, M. (1997, Winter). Assessment as a lever in education reform. *Phi Kappa Phi: National Forum, 77*(1), 27- 30.

Diez, M. (1998). The role of standards and assessment: A dialogue. In M. Diez (Ed.), *Changing the practice of teacher education: Standards and assessment as a lever for change*. Washington, DC: AACTE.

Diez, M., & Hass, J. (1997). No more piecemeal reform: Using performance-based approaches to rethink teacher education. *Action in Teacher education, XIX*(2), 17-26.

Digest of Educational Statistics. (1997). *Elementary and secondary: Regulations* (NCES 98-015). Washington, DC: National Center for Education Statistics.

Dill, W. (1998, July/August). Specialized accreditation: An idea whose time has come? Or gone? *Change*, 18-25.

Dunkin, M., & Biddle, B. (Eds.). (1974). *The study of teaching*. New York, NY: Holt, Rinehart, and Winston.

Earley, P. (2000). Finding the culprit: Federal policy and teacher education. *Educational Policy, 14*(1), 25- 39.

Education Commission of the States. (2000). *In pursuit of quality teaching: Five key strategies for policymakers*. Denver, CO: Author.

Elmore, R., & Burney, D. (1997, August). *Investing in teacher learning: Staff development and instructional improvement in Community School District #2, New York City*. New York, NY: National Commission on Teaching & America's Future.

Engel, M. (2000). *The struggle for control of public education: Market ideology vs. democratic values*. Philadelphia, PA: Temple University Press.

Erickson, F. (1986). Qualitative methods on research on teaching. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 119-161). New York, NY: Macmillan.

Evertson, C., Hawley, W., & Zlotnik, M. (1985). Making a difference in educational quality through teacher education. *Journal of Teacher Education, 36*, 2-10.

Farkas, S., & Johnson, J. (1997). *Different drummers: How teachers of teachers view public education*. New York, NY: Public Agenda.

Fenstermacher, G. (1994). The knower and the known: The nature of knowledge in research on teaching. In L. Darling- Hammond (Ed.), *Review of research in education* (Vol. 20, pp. 3-56). Washington, DC: American Educational Research Association.

Ferguson, P., & Womack, S. (1993). The impact of subject matter and education coursework on teacher performance. *Journal of Teacher Education, 44*(1), 55-63.

Finn, C., Kanstoroom, M., & Petrilli, M. (1999, November). *The quest for better teachers: Grading the states*. Washington, DC: Thomas B. Fordham Foundation.

Finn, C., & Petrilli, M. (Eds.). (2000, January). *The state of state standards 2000*. Washington, DC: Thomas B. Fordham Foundation.

Flippo, R. (1986). Teacher certification testing: Perspectives and issues. *Journal of Teacher Education, 37*(2), 2-9.

Floden, R., & Buchman, M. (1990). Philosophical inquiry in teacher education. In W. Houston (Ed.), *Handbook of research on teacher education* (pp. 42-58). New York, NY: Macmillan.

Fordham Foundation. (1999). *Manifesto: A policy statement*. New York, NY: Thomas B. Fordham Foundation.

Freire, P. (1970). *Pedagogy of the oppressed* (M. B. Ramos, Trans.). New York, NY: Seabury Press.

Gage, N. (1972). *Teacher effectiveness and teacher education: The search for a scientific basis*. Palo Alto, CA: Pacific Books.

Garcia, P. (1986). The impact of national testing on ethnic minorities: With proposed solutions. *Journal of Negro Education, 55*(3), 347-357.

Gardner, W. (1989). Preface. In M. Reynolds (Ed.), *Knowledge base for the beginning teacher*. New York, NY: Pergamon Press.

Gay, G. (2000). *Culturally responsive teaching: Theory, research and practice*. New York, NY: Teachers College Press.

Gitomer, D., & Latham, A. (2000). Generalizations in teacher education: Seductive and misleading. *Journal of Teacher Education, 51*(3), 215-220.

Goldhaber, D., & Brewer, D. (1999). Teacher licensing and student achievement. In M. Kanstoroom & C. Finn (Eds.), *Better teaches, better schools*. Washington, DC: Thomas B. Fordham Foundation.

Graham, P., Lyman, R., & Trow, M. (1998). *Accountability of colleges and universities: An essay*. New York, NY: Columbia University Press.

Griffin, G. (1999). Changes in teacher education: Looking to the future. In G. Griffin (Ed.), *The education of teachers: Ninety-eighth yearbook of the National Society for the Study of Education* (pp. 1-17). Chicago, IL: University of Chicago Press.

Grimmett, P., & MacKinnon, A. (1992). Craft knowledge and the education of teachers. *Review of research in education, 18*, 385-456.

Grossman, P. (1990). Mapping back: The role of subject- specific teacher education. In P. Grossman (Ed.), *The making of a teacher: Teacher knowledge and teacher education*. New York, NY: Teachers College Press.

Haberman, M. (1985). Does teacher education make a difference: A review of comparisons between liberal arts and teacher education graduates. *Journal of Thought, 20*(2), 25-34.

Haertel, E. (1999, May). Performance assessment and education reform. *Phi Delta Kappan, 80*(9), 662- 666.

Hawley, W., & Valli, L. (1999). The essentials of effective professional development: A new consensus. In L. Darling- Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 127- 150). San Francisco, CA: Jossey Bass.

Heaton, R. M., & Lambert, M. (1993). Learning to hear voices: Inventing a new pedagogy of teacher education. In D. K. Cohen & M. W. McLaughlin & J. E. Talbert (Eds.), *Teaching for understanding: Challenges for policy and practice* (pp. 43-83). San Francisco, CA: Jossey-Bass Publishers.

Hiebert, J. (1999). Relationships between research and the NCTM standards. *Journal for Research in Mathematics Education, 30*(1), 3-19.

Holmes Group. (1986). *Teachers for tomorrow's schools*. East Lansing, MI: Author.

Holmes Group. (1996). *Tomorrow's schools of education*. East Lansing, MI: Author.

hooks, b. (1994). *Teaching to transgress: Education as the practice of freedom:* New York, NY: Routledge.

Houston, R. W. (Ed.). (1990). *Handbook of research on teacher education* (1st ed.). New York, NY: Macmillan.

Hudson, A., & Lambert, D. (Eds.). (1997). *Exploring futures in initial teacher education*. London: University of London: Institute of Education.

Imig, D. (2000, October). Personal Communication.

Interstate New Teacher Assessment and Support Consortium. (1992). *Model standards for beginning teacher licensure and development: A resource for state dialogue*. Washington, DC: Council of Chief State School Offices.

Irvine, J. (1990). *Black students and school failure: Policies, practice and prescriptions*. New York, NY: Greenwood Press.

Irvine, J., & Fraser, J. (1998, May 13). *Warm demanders*, [online website]. http://www.edweek.org/ew/1998/35irvine.h17

Irvine, J., & York, D. E. (1995). Learning styles and culturally diverse students: A literature review. In J. A. Banks & C. A. M. Banks (Eds.), *Handbook of research on multicultural education*. New York, NY: Macmillan.

Jacobson, S., Emihovich, C., Helfrich, J., Paetrie, H., & Stevenson, H. (1998). *Transforming schools and schools of education: A new vision for preparing educators*. Thousand Oaks, CA: Corwin Press.

Kanstoroom, M., & Finn, C. (1999, July). *Better teachers, better schools*. Washington, DC: Thomas B. Fordham Foundation.

Kaplan, L., & Edelfelt, R. (Eds.). (1996). *Teachers for a new millennium: Aligning development, national goals, and high standards for all students*. Thousand Oaks, CA: Corwin Press.

Kennedy, M. (Ed.). (1991). *Teaching academic subjects to diverse learners*. New York, NY: Teachers College Press.

King, J. E., Sleeter, C. E., Gutierrez, W., New, C. A., & Takata, S. R. (1996). Race and education: In what ways does race affect the educational process? In J. L. Kincheloe & S. Steinberg (Eds.), *Thirteen questions: Reframing education's conversation* (2nd ed., pp. 157-190). New York, NY: Peter Lang Publishers.

Klagholtz, L. (2000). *Growing better teachers in the garden state: New Jersey's "alternate route" to teacher certification*, [online website].
http://www.edexcellence.net/library/newjersey/new9620jersey. html.

Klausmeier, R. (1990). Four decades of calls for reform of teacher education: The 1950s through the 1980s. *Teacher Education Quarterly, 17*(4), 23-64.

Labaree, D. (1997). Public goods, private goods: The American struggle over educational goals. *American Educational Research Journal, 34*(1), 39-81.

Ladson-Billings, G. (1994). *The dream keepers: Successful teachers of African-American children*. San Francisco, CA: Jossey Bass.

Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal, 32*(3), 465-491.

Ladson-Billings, G. (1998). Teaching in dangerous times: Culturally relevant approaches to teacher assessment. *Journal of Negro Education, 67*(3), 255-267.

Ladson-Billings, G. (1999a). Preparing teachers for diverse student populations: A critical race theory perspective. In A. Iran-Nejad & D. Pearson (Eds.), *Review of research in education* (Vol. 24, pp. 211-248). Washington, DC: American Educational Research Association.

Ladson-Billings, G. (1999b). Preparing teachers for diversity: Historical perspectives, current trends, and future directions. In L. Darling-Hammond & G. Sykes (Eds.), *Learning as the teaching profession: Handbook of policy and practice* (pp. 86-124). San Francisco, CA: Jossey Bass.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Lampert, M., & Ball, D. (1998). *Teaching, multimedia, and mathematics: Investigations of real practice*. New York, NY: Teachers College Press.

Lampert, M., & Ball, D. (1999). Aligning teacher education with contemporary K-12 reform visions. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 33-53). San Francisco, CA: Jossey Bass.

Levine, M., & Trachtman, R. (Eds.). (1997). *Making professional development schools work: Politics, practice, and policy*. New York, NY: Teachers College Press.

Lieberman, A., & Miller, L. (1994). Problems and possibilities of institutionalizing teacher research. In S. Hollingsworth & H. Socket (Eds.), *Teacher research and educational reform* (pp. 204-220). Chicago, IL: The University of Chicago Press.

Little, J. (1993). Professional community in comprehensive high schools: The two worlds of academic and vocational teachers. In J. Little & M. McLaughlin (Eds.), *Teacher's work: Individuals, colleagues, and contexts* (pp. 137- 163). New York, NY: Teachers College Press.

Loucks-Horsley, S. (1995). Professional development and the learner centered school. *Theory into Practice, 34*(4), 265-271.

Lucas, C. (1999). *Teacher education in America: Reform agendas for the twenty-first century*. New York, NY: St. Martin's Press.

Luna, C., Solsken, J., & Kutz, E. (2000). Defining literacy: Lessons from high-stakes teacher Testing. *Journal of Teacher Education, 51* (4), 276-288.

Lytle, S., & Cochran-Smith, M. (1992). Teacher research as a way of knowing. *Harvard Educational Review, 62*(4), 447-474.

Madaus, G. (1993). National testing system: Manna from above? A historical perspective. *Educational Assessment, 1*, 9-26.

Madaus, G., & O'Dwyer, L. (1999). A short history of performance assessment: Lessons learned. *Phi Delta Kappan, 80*(9), 688-695.

Maggio, R. (Ed.). (1997). *Quotations on education*. Paramus, NJ: Prentice Hall.

Maimon, J. (1999). Little pigs and big ideas: Blowing down the 'at-risk' straw house. *International Journal of Leadership in Education: Theory and Practice, 2*(3), 191- 206.

McCarthy, C. (1993). Multicultural approaches to racial inequality in the United States. In L. A. Castenell & W. F. Pinar (Eds.), *Understanding curriculum as racial text*. Albany, NY: SUNY Press.

McConney, A., Schalock, M., & Schalock, D. (1998). Focusing improvement and quality assurance: Work samples as authentic performance measures of prospective teachers' effectiveness. *Journal of Personnel Evaluation in Education, 12*(3), 237-246.

McLaughlin, M., & Talbert, J. (1993). Introduction: New visions of teaching. In D. K. Cohen & M. W. McLaughlin & J. E. Talbert (Eds.), *Teaching for understanding: Challenges for policy and practice* (pp. 1-12). San Francisco, CA: Jossey-Bass Publishers.

Melnick, S., & Pullin, D. (2000). 'Can you take dictation?' Prescribing teacher quality through testing. *Journal of Teacher Education, 51*(4), 262-275.

Millman, J. (1997).(Ed.), *Grading teachers, grading schools* Thousand Oaks, CA: Corwin Press.

Mosle, S. (1996, October 27). The answer is national standards. *New York Times*.

Murray, F. (Ed.). (1996). *The teacher educator's handbook: Building a knowledge base for the preparation of teachers*. Washington, DC: AACTE.

Murray, F. (2000). The role of accreditation reform in teacher education. *Educational Policy, 14*(1), 40- 60.

National Board for Professional Teaching Standards. (1994). *Standards*. Detroit, MI: NBPTS.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U.S. Government Printing Office.

National Commission on Teaching & America's Future. (1996, September). *What matters most: Teaching for America's future*. New York, NY: Teachers College, Columbia University.

National Commission on Teaching and America's Future. (1997, November). *Doing what matters most: Investing in quality teaching*. New York, NY: National Commission on Teaching and America's Future.

National Council for the Accreditation of Teacher Educators. (1999). *Proposed NCATE 2000 unit standards*, [online website]. http://www.ncate.org.

Nieto, S. (1996). *Affirming diversity: The sociopolitical context of* *multicultural education*. New York, NY: Longman.

Noffke, S. (1997). Professional, personal, and political dimensions of action research. *Review of research in education, 22*, 305-343.

Patterson, R., Michelli, N., & Pacheco, A. (1999). *Centers of pedagogy: New structures for educational renewal*. San Francisco, CA: Jossey Bass.

Ravitch, D. (2000). *Left back: A century of failed school reforms*. New York, NY: Simon & Schuster.

Reynolds, M. C. (Ed.). (1989). *Knowledge base for the beginning teacher*. Oxford: Pergamon Press.

Robertson, H. (1998). Public education in a corporate- dominated culture. In A. Hargreaves & A. Lieberman & M. Fullan & D. Hopkins (Eds.), *International handbook of educational change* (pp. 396-417). Boston, MA: Kluwer Academic Publishers.

Roth, R. (1996). Standards for certification, licensure, and accreditation. In J. Sikula & T. J. Buttery & E. Guyton (Eds.), *Handbook of research on teacher education* (2nd ed., pp. 242-278). New York, NY: Macmillan.

Sanders, W., & Horn, S. (1994). The Tennessee Value-Added Assessment System (TVAAS): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education, 8*, 299-311.

Sanders, W., & Horn, S. (1998). Research findings from the Tennessee Value-Added Assessment System (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education, 12*(3), 247-256.

Scannell, M., & Metcalf, P. (2000). Autonomous boards and standards-based teacher development. *Educational Policy, 14*(1), 61-76.

Schalock, D., & Imig, D. (2000). *Shulman's union of insufficiencies +7: New dimensions of accountability for teachers and teacher educators*. Washington, DC: American Association of Colleges for Teacher Education.

Schalock, D., & Myton, D. (1988). A new paradigm for teacher licensure: Oregon's demand for evidence of success in fostering learning. *Journal of Teacher Education, 39*(6), 27-32.

Schalock, D., Schalock, M., & Girod, G. (1997). Reflections on comments by Airasian and Stufflebeam. In J. Millman (Ed.), *Grading* *teachers, grading schools* (pp. 62-71). Thousand Oaks, CA: Corwin Press.

Schalock, D., Schalock, M., & Myton, D. (1998, February). Effectiveness —along with quality —should be the focus. *Phi Delta Kappan*, 468-470.

Schon, D. (1983). *The reflective practitioner*. San Francisco, CA: Jossey-Bass.

Schon, D. (1987). *Educating the reflective practitioner*. San Francisco, CA: Jossey-Bass.

Schrag, P. (1999, July/August). Who will teach the teachers? *University Bulletin*.

Shulman, J. (1992). *Case methods in teacher education*. New York, NY: Teachers College Press.

Shulman, L. (1986). Paradigms and research programs in the study of teaching: A contemporary perspective. In M. C. Wittrock (Ed.), *Handbook of research on teaching*.

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 51*, 1- 22.

Sikula, J., Buttery, T. J., & Guyton, E. (Eds.). (1996). *Handbook of research on teacher education* (2nd ed.). New York, NY: Macmillan.

Sirotnick, K., & Goodlad, J. (Eds.). (1988). *School- university partnerships in action*. New York, NY: Teachers College Press.

Sleeter, C. (1992). Restructuring schools for multicultural education. *Journal of Teacher Education, 43*(2), 141- 148.

Sleeter, C., & Grant, C. (1987). An analysis of multicultural education in the United States. *Harvard Educational Review, 57*(4), 421-444.

Sleeter, C. E., & McLaren, P. L. (1995). *Multicultural education, critical pedagogy, and the politics of difference*. Albany, NY: State University of New York Press.

Smith, G. (1984). The critical issue of excellence and equity in competency testing. *Journal of Teacher Education, 35*(2), 6-9.

Smith, M. L., Heinecke, W., & Noble, A. (1999, Winter). Assessment policy and political spectacle. *Teachers College Record, 101*(2), 157-191.

Stein, M., Smith, M., & Silver, E. (1999, Fall). The development of professional developers: Learning to assist teachers in new settings in new ways. *Harvard Educational Review, 69*(3), 237-269.

Stotsky, S. (1999). *Losing our language: How multicultural classroom instruction is undermining our children's ability to read, write, and reason*. New York, NY: Free Press.

Sykes, G. (1999). Introduction: Teaching as the learning profession. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. xv-xxiv). San Francisco, CA: Jossey Bass.

Symms Gallagher, K., & Bailey, J. (2000). Introduction to the politics of teacher preparation reform. *Educational Policy, 14*(1), 6-10.

Symms Gallagher, K., & Bailey, J. (2000). The politics of teacher education reform: Strategic philanthopy and public policy making. *Educational Policy, 14*(1), 11-24.d

Tatum, B. (1992). Talking about race, learning about racism. The applications of racial identity development theory. *Harvard Educational Review, 62*(1), 1-24.

Teacher Education Accreditation Council. (1999). *Accreditation principles of the TEAC*, [online website]. http://www.teac.org/

Theissen, D. (2000). Developing knowledge for preparing teachers: Redefining the role of schools of education. *Educational Policy, 14*(1), 129-144.

Urban, W. (1990). Historical studies of teacher education. In W. Houston (Ed.), *Handbook of research on teacher education* (pp. 59-71). New York, NY: Macmillan.

U. S. Department of Education. (1998, September). *Promising practices: New ways to improve teacher quality*, [online website]. http://www.ed.gov/pubs/PromPractice/index.html

U. S. Department of Education. (1999). *Investing in learning: A policy statement with recommendations on research in education by the National Educational Research Policy and Priorities Board*, [online website]. http://www.ed.gov/pubs/InvestInLearning.

U. S. Department of Education. (2000, February 1). *Update on federal reporting requirements for teacher preparation institutions and programs*. Washington, DC: U. S. Department of Education.

Weiner, L. (1989). Asking the right questions: An analysis framework for reform of urban teacher education. *The Urban Review, 21*(3), 151-

161.

Whitty, G., Power, S., & Halpin, D. (1998). *Devolution and choice in education: The school, the state, and the market*. Melbourne, Australia: Australian Council for Educational Research.

Wilson, S., & Ball, D. (1996). Helping teachers meet the standards: New challenges for teacher educators. *Elementary School Journal, 97* (2), 121-138.

Wise, A. (1988). *Impacts of teacher testing: State educational governance through standard-setting* (NIE-G- 83-0023). Santa Monica, CA: Rand Corporation.

Wise, A. (1996, November). Building a system of quality assurance for the teaching profession: Moving into the 21st century. *Phi Delta Kappan, 78*(3), 191-192.

Wise, A. (1999). Technos interview: Arthur E. Wise. *Technos, 8*(3), 6-11.

Yarger, S., & Smith, P. (1990). Issues in research on teacher education. In W. Houston (Ed.), *Handbook of research on teacher education* (pp. 25-41). New York, NY: Macmillan.

Yinger, R. (1999). The role of standards in teaching and teacher education. In G. Griffin (Ed.), *The education of teachers: Ninety-eighth yearbook of the National Society for the Study of Education* (pp. 85-113). Chicago, IL: University of Chicago Press.

Yinger, R., & Hendricks-Lee, M. (2000). The language of standards and teacher education reform. *Educational Policy, 14*(1), 94-106.

Zeichner, K. (1988). *Understanding the character and quality of the academic and professional components of teacher education* (ERIC Reproduction Services, ED301537).

Zeichner, K. (2000). Ability-based teacher education: Elementary education at Alverno. In L. Darling-Hammond (Ed.), *Studies of excellence in teachers*. Washington, DC: National Commission on Teaching and America's Future.

Zeichner, K., & Liston, D. (1987). Teaching student teachers to reflect. *Harvard Educational Review, 57*, 1-22.

About the Author

Marilyn Cochran-Smith
Professor of Education

Boston College
Lynch School of Education
Campion Hall 113
Chestnut Hill, MA 02467
617-552-4180
Email: cochrans@bc.edu

Marilyn Cochran-Smith is Professor of Education and Director of the Doctoral Program in Curriculum & Instruction at the Lynch School of Education at Boston College. She is immediate past Vice President of AERA for Division K (Teaching and Teacher Education, 1998-2000). Cochran-Smith is the Editor of the *Journal of Teacher Education* and co-chair (with Ken Zeichner) of the AERA National Consensus Panel on Teacher Education as well as co-editor (with Susan Lytle) of the Teachers College Press book series on *Practitioner Inquiry*. She is also a member of the advisory board for the Carnegie Foundation for the Advancement of Teaching's CASTL project (Carnegie Academy for the Scholarship of Teaching and Learning) as well as the new Teacher Education Committee of the National Academy of Education. She has written extensively about inquiry-based teacher education, race and diversity issues in teacher education,and teacher research as both professional development and knowledge generation.

Copyright 2001 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin
Greg Camilli
Rutgers University
John Covaleskie
Northern Michigan University
Alan Davis
University of Colorado, Denver
Sherman Dorn
University of South Florida
Mark E. Fetler
California Commission on Teacher Credentialing
Richard Garlikov
hmwkhelp@scott.net
Thomas F. Green
Syracuse University
Alison I. Griffith
York University
Arlen Gullickson
Western Michigan University
Ernest R. House
University of Colorado
Aimee Howley
Ohio University

Craig B. Howley
Appalachia Educational Laboratory

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

William Hunter
University of Calgary

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

David D. Williams
Brigham Young University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas
(México)

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)

Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

other vols. | abstracts | editors | board | submit | comment | subscribe | search